THE UNIVERSITY
of EDINBURGH

Evolutionary history of a clone of *Staphylococcus aureus* infecting multiple host-species.

Laura Spoor

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2014

I declare that I have composed this thesis, and that the research presented within is my own work, except where stated otherwise. This thesis has not been submitted for any other degree or professional qualification.

Laura Spoor November 2014

Acknowledgements

I would like to thank my supervisor Professor Ross Fitzgerald for his invaluable support, expertise and guidance throughout my PhD. Thanks also to my second supervisor Dr Geoff Faulkner for assisting me with python coding problems.

I am very grateful to all "LBEPers" old and new alike, including Gill, Bethan, Robyn, Anne, Mark, Helen, Emma, Helene, Kaj, Amy, Stephen, Emily, Cheryl, Rodrigo, Mariya, Gordon, Chriselle and Melissa. Special Thanks to Robyn Cartwright for sharing her fantastic lab expertise, excellent baking skills, and support inside and outside the lab. I am very grateful to Paul McAdam who collaborated with me, and helped me hugely on the road to becoming a bioinformatician! Thanks to Emily Richardson for her help and support, and Amy Richards for her laboratory collaboration. Thanks also to Louise Welch, and all those at Roslin who have made my time here so fantastic.

I am grateful to Dr Andrew Rambaut for his BEAST advice, and I am grateful to Dr Lucy Weinert at the University of Cambridge and Gina McAllister for their collaborations with me. Thanks to GenePool and ARK Genomics (University of Edinburgh) for sequencing services, and for my support from the Swan-Riddell Ker memorial scholarship and the Royal (Dick) School of Veterinary studies. Thanks to A. Battisti, E. Smith, L. Fox, J. Kruze, R. de Campo, J. Lindsay, H. de Lencastre, R. Miller, S. Monecke, F. Layer, B. Strommenger, W. Witte, M. Sudagidan, K. R. N dos Santos, R. Ruimy, H. Westh, A. Kearns, M. Ellington, and E. Feil for kindly sending me strains.

Last but not least a huge Thank you to Al Vaughan for his love and fantastic support, my family Judith, David and Helen, and to Lynne, Richard, Sam, Dennis, Eileen, Jenny, Baxter and my extended RAF family, all of whom have provided continuous support and encouragement that has been an inspiration throughout.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full description
α	Alpha
β	Beta
°C	Degrees Celsius
AP buffer	Antarctic phosphatase buffer
BLASTN/BLASTX	Basic local alignment search tool
bp	Base pair
BSE	Bovine spongiform encephalopathy
CA-MRSA	Community-associated methicillin-resistant *Staphylococcus aureus*
CC	Clonal complex
CDS	Protein coding sequence
CI	Consistency index
CWA	Cell wall-associated
dH20	Distilled water
DLV	Double-locus variant
DMSO	Dimethyl sulfoxide
dN	Nonsynonymous mutation
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate mix (dATP, dCTP, dGTP, dTTP)
dS	Synonymous mutation
EARS-Net	European Antimicrobial Resistance Surveillance Network
EMRSA	Epidemic methicillin-resistant *Staphylococcus aureus*
ExoAP mix	Exonuclease I and Antarctic Phosphatase mix
gDNA	Genomic deoxyribonucleic acid

Abbreviation	Full description
h	Hour
HA-MRSA	Hospital-associated methicillin-resistant *Staphylococcus aureus*
HGT	Horizontal gene transfer
HPD	Highest posterior density
IEC	Immune evasion cluster
Ig	Immunoglobulin
Indels	Small insertions and deletions
IS	Insertion sequence
kbp	Kilo-base
LA-MRSA	Livestock-associated methicillin-resistant *Staphylococcus aureus*
MEGA	Molecular Evolutionary Genetics Analysis
mg	Milligrams
$MgCl_2$	Magnesium chloride
MGE	Mobile Genetic Element
MIC	Minimum inhibitory concentration
min	Minutes
ml	Millilitres
MLEE	Multi-locus enzyme electrophoresis
MLST	Multi-locus sequence type
MLVA	Multi-locus variable number of tandem repeat analysis
mM	Millimolar
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSSA	Methicillin-susceptible *Staphylococcus aureus*
NCBI	National Centre for Biotechnology Information

Abbreviation	Full description
ng	Nanograms
nmol	Nanomolar
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PIA	Polysaccharide intercellular adhesin
PSA	Proportion of shared ancestry tree
PVL	Panton-Valentin Leukocidin
RAST	Rapid Annotation using Subsystem Technology
RDP	Recombination detection program
RM system	Restriction modification system
s	Seconds
SaPI	Staphylococcal pathogenicity island
SaPIbov	Bovine staphylococcal pathogenicity island
SARS	Severe acute respiratory syndrome
SCC	Somatic cell counts
SCC	Staphylococcal cassette chromosome
SCC*mec*	Staphylococcal cassette chromosome *mec*
SLV	Single-locus variant
SNPs	Single nucleotide polymorphisms
SRA	Sequence read archive
ST	Sequence type
SWP	South-west pacific
T4SS	Type IV secretion system
TBE	Tris-borate/EDTA
TSA	Tryptone soya agar
TSB	Tryptone soya broth
TSST-1	Toxic shock syndrome toxin-1

Abbreviation	Full description
U	Units
µg	Micrograms
µl	Microlitres
v	Voltage
v/v	Volume for volume
vWbp	Von Willebrand factor binding protein
WGS	Whole genome sequencing

Abstract

Staphylococcus aureus is an important opportunistic pathogen in humans and animals. In humans, there has been an increase in community-associated methicillin-resistant *Staphylococcus aureus* (MRSA) causing disease in healthy humans. The exact evolutionary origins and basis for its recent expansion are not yet clear. In livestock, *S. aureus* is an important cause of diseases of welfare and economic importance, including bovine mastitis.

Molecular typing studies demonstrate that natural populations of *S. aureus* are highly clonal and largely adapted to a specific host, however there are some lineages that colonise multiple host species. In particular, clonal complex 97 (CC97) is a dominant bovine mastitis-associated lineage which has been isolated from other animal species, and more recently there are increasing reports of CC97 *S. aureus* from human infections worldwide. The basis for this wide host tropism is currently unknown. In order to investigate the evolutionary origins of *S. aureus* CC97, 43 strains were selected for whole genome sequencing, isolated from humans, cattle and pigs, from 18 different countries on 4 continents, ranging from 1956 to 2012. Phylogenetic analysis using high quality core genome single nucleotide polymorphisms (SNPs) resolved the single CC97 lineage into host-adapted sublineages, which were likely the result of 2 independent livestock-to-human host jumps estimated to have occurred at least 40 years ago. One of the human sublineages consisted of strains from 4 continents indicating global dissemination since the host jump occurred.

In order to investigate the genetic basis for human host adaptation in *S. aureus* CC97, comparative genomic analysis of mobile genetic elements, nonsynonymous SNPs and small insertions and deletions was performed. Of note, independent acquisitions of genetic elements encoding antimicrobial resistance and specific mediators of human innate immune evasion were identified in the human-adapted *S. aureus* CC97 strains. These data are consistent with an important role for mobile genetic elements in the host adaptive evolution of *S. aureus* CC97.

Also in the current study, a bovine-associated single locus variant of ST97 (ST71) was identified as a phylogenetic outgroup relative to all other *S. aureus* CC97 strains examined. Comparative genomic analysis of ST71 strains with representative bovine

ST97 strains indicate that ST71 has a mosaic genome. A large region spanning the origin of replication demonstrated closest homology to non-CC97 ruminant-associated genotypes, with the remainder of the genome consistent with an ancestral ST97 genetic background. Recombination detection analysis predicts that one or more large-scale recombination events have occurred in the region that spans the origin of replication, resulting in variation in gene content between ST71 and ST97. The data highlight the potential role of homologous recombination in rapidly generating genomic diversity that might alter the phenotype of strains in the ecological niche of the bovine mammary gland.

Overall, the study reveals the evolutionary history of a major pathogenic clone of *S. aureus* affecting multiple host species, and identifies the genetic events which have contributed to its success.

1. Introduction

1.1 The human-animal interface in infectious disease

As a result of the domestication of livestock species such as cows, goats, sheep and pigs approximately 11 000 years ago, animals and humans have remained in close contact for thousands of years, with populations continually expanding and agricultural systems intensifying (Zeder, 2008). This has provided ample transmission opportunities for pathogens, with animals, including livestock and wildlife, often implicated as either the origin or reservoir for several zoonotic diseases responsible for causing major human morbidity and mortality (Wolfe, Dunavan, and Diamond, 2007; Woolhouse and Gaunt, 2007). It has been estimated that approximately 80 % of emerging pathogens since 1980 are associated with an animal reservoir or origin, although almost 75 % of these are viruses (Woolhouse and Gaunt, 2007).

For example, highly pathogenic avian influenza (HPAI) H5N1 is now endemic in poultry in Asia, with a reservoir in wildfowl providing re-infection opportunities, resulting in repeated zoonotic transmission events to humans in direct contact with poultry (Li et al., 2004; Chotpitayasunond et al., 2005; Tran et al., 2004). The severe acute respiratory syndrome (SARS) epidemic caused by a coronavirus (SARS-CoV) in 2002-2003, resulting in over 8000 human cases, was attributed to zoonotic infection from palm civets and racoon dogs in the wet markets of China (Guan et al., 2003; Song et al., 2005). However, other studies implicate bats as the origin, with civets as a possible intermediate or ancilliary hosts (Ge et al., 2013; Lau et al., 2005; Li et al., 2005). In other cases, human pathogens appear to have evolved from an ancestor originally found in animals a long time ago, such as the Morbillivirus causing measles in humans, which appears to have diverged from rinderpest virus in cattle (Furuse, Suzuki, and Oshitani, 2010).

Although the reason behind the over-representation of viruses is not known (Wolfe, Dunavan, and Diamond, 2007), there are some important human bacterial pathogens

that have been predicted as having animal-associated origins. For example, the causative agent of the human plague, *Yersinia pestis*, is thought to have originated in a rodent host (Cui et al., 2013). The closest relatives of *Corynebacterium diphtheriae*, the causative agent of diphtheria, are *C. pseudotuberculosis* or *C. ulcerans*, both of which colonise domestic herbivorous livestock and, as such, livestock have been implicated as a probable origin by Wolfe *et al* (Martínet al., 2003; Wolfe, Dunavan, and Diamond, 2007). Although *Mycobacterium tuberculosis*, the pathogen responsible for causing human tuberculosis (TB) was previously thought to have originated as a zoonosis from the ruminant pathogen *M. bovis* (Stead et al., 1995), more recently the consensus view is that animal-associated species of Mycobacteria including *M. bovis*, are derived from a human-associated ancestral relative of *M. tuberculosis* (Brosch et al., 2002; Garnier et al., 2003; Hershberg et al., 2008). However, overall, there is a lack of strong phylogenetic evidence, leading to uncertainty regarding the true evolutionary origins of these bacterial pathogens.

The potential for the emergence or re-emergence of zoonotic pathogens is a continual threat to global human health and likely to be affected by factors associated with modern civilisation, including the ever-expanding global population, modern air travel, climate change and the intensification of agricultural practices, all of which may affect the evolution of pathogens or transmission dynamics at the human-animal interface. While some of the wide-ranging ecological factors are reviewed by Jones *et al* (Jones et al., 2013), some examples are listed here that highlight how the intensification of farming practices has altered pathogen transmission dynamics and the emergence of diseases of importance to human and veterinary medicine.

The use of antimicrobials at subtherapeutic doses in animal feed for the promotion of optimal growth of livestock has been linked to selection for resistant bacteria that might complicate future treatment of bacterial infections in human and veterinary medicine (Bager et al., 1997; Bates, Jordens, and Griffiths, 1994). For example, the use of Avoparcin as a growth promoter in feed premix has been implicated in

causing cross-resistance of *Enterococci* to other clinically-important glycopeptides used to treat human bacterial infections, such as vancomycin, although an EU-wide ban on the use of antimicrobials as growth promoters came into force in 1996 (Bager et al., 1997; Bates et al., 1994). However, in the United States, antimicrobial use at therapeutic and subtherapeutic dose rates (growth promoters) in agriculture is currently unregulated and widespread (Food and Drug Administration, 2010).

The bovine spongiform encephalopathy (BSE) crisis in cattle and associated human variant Creutzfeldt Jakob disease (vCJD) arose as a result of recycling cheap animal protein in the form of meat and bone meal that was contaminated with prion proteins, which was then ingested by cows, causing BSE, and subsequently vCJD in humans after ingesting contaminated cow meat (Smith, 2003).

Epidemiological analysis of a human outbreak of Nipah Virus encephalitis in Malaysia and Singapore found evidence for the intensification of livestock production as a driver for the emergence of Nipah Virus between the primary host, bats and the intermediate host pigs, which provided increased infection opportunities subsequently in humans (Pulliam et al., 2012). Data on management practices from the index farm along with epidemiological knowledge on bat populations, indicate that production methods such as the dual use of land for mango tree production promoted repeated introduction of the Nipah virus from bats that use the trees to roost to adjacent pig sheds. This enabled enzootic persistence in the immunologically primed pig herd, and spread among pigs through transport to local farms, resulting in further human infections through repeated contact with persistently infected pigs (Pulliam et al., 2012).

Taken together, these factors highlight the importance of continued surveillance at the interface of human and animal contact, which has been recognised through such broad concepts as "One Health", promoting a multi-disciplinary collaborative

approach to disease surveillance and control in humans and animals at the local, national, and global level (Alder and Easton, 2005; Coker et al., 2011). One such area for research focus is to investigate pathogens known to infect multiple host species in order to understand how cross-species transmission occurs, which may inform biosecurity control measures or therapeutic interventions such as vaccination.

1.2 *S. aureus* in humans

1.2.1 Clinical significance

Staphylococcus aureus is a Gram positive coccus-shaped bacterium that exhibits tremendous diversity in host tropism, and in the range and severity of clinical manifestations that it is associated with (Lowy, 1998; Smyth et al., 2009). In humans, *S. aureus* primarily colonises the anterior nares, with longitudinal studies estimating that approximately 20 % to 30 % of people are persistent carriers of *S. aureus* (Kluytmans, Van Belkum, and Verbrugh, 1997; Williams, 1963). However, persistent carriage has been shown to be a risk factor for clinical *S. aureus* infections (von Eiff et al., 2001), and *S. aureus* is an opportunistic pathogen which can cause a myriad of clinical manifestations, ranging from mild skin infections through to life-threatening illness such as septicaemia, endocarditis and pneumonia (Frazee et al., 2005; Lowy, 1998). *S. aureus* produces a variety of secreted virulence factors including toxins, which can cause toxic shock syndrome, scarlet fever, scalded skin syndrome and food poisoning (Kapral and Miller, 1971; Le Loir, Baron, and Gautier, 2003; Parsonnet et al., 2010).

1.2.2 Hospital-associated methicillin-resistant *S. aureus* (HA-MRSA)

S. aureus is one of the major causes of nosocomial infections worldwide (Diekema et al., 2001). The first report of methicillin-resistant *S. aureus* (MRSA) was in the UK in 1961, which was shortly after the introduction of methicillin into clinical use (Jevons, 1961). The rate of MRSA infections due to HA-MRSA has increased to the point at which it is now endemic within healthcare settings in many countries worldwide (Klevens et al., 2007). In Europe, HA-MRSA infection rates vary hugely, ranging from an extremely low prevalence of less than 1 % in northern European countries such as Sweden, to higher infection rates such as 14 % in the UK, with the highest rates of 53 % observed in Portugal and Romania (EARS-Net, 2012). The United States has high rates of MRSA infection, with one study estimating 31.8 cases per 100 000, and a mortality rate estimated at 6.3 per 100 000 (Klevens et al., 2007). In some countries, the proportion of nosocomial infections due to HA-MRSA is high, with infection rates of 86.5 % in Sri Lanka, 77.6 % in Korea and 74.1 % in Vietnam reported (Song et al., 2011). However, in recent years, several European countries including the UK have noticed a significant overall decline in MRSA infection rates (EARS-Net, 2012). For example, one of the major epidemic MRSA strains in the UK, known as EMRSA-16 has declined in prevalence from 21 % to 9 % between 2001 and 2007 (Ellington et al., 2010). However, longitudinal studies indicate that over time, the dominant clone circulating in a hospital can change, such that while some clones of MRSA decrease in prevalence, others may increase (Knight et al., 2012). For example, in a study of HA-MRSA over a 10 year period in a UK hospital, the predominant clone was found to switch from EMRSA-16 (CC30) to EMRSA-15 (CC22) after 2003 (Knight et al., 2012). A clone known as CC239 was also detected in 2003 and increased in prevalence initially, but had almost disappeared entirely by 2006 (Knight et al., 2012).

Some of the risk factors associated with acquiring HA-MRSA include indwelling intravenous devices, previous treatment with antimicrobials, previous hospitalisation, surgery and length of stay prior to MRSA infection (Coello, et al., 1997; Dziekan et

al., 2000; Graffunder, 2002). In addition to the fact that HA-MRSA affects already-compromised individuals, treatment options can also be complicated by resistance to multiple classes of antimicrobial, such as that described in the major HA-MRSA strain, ST239 (Ko et al., 2005; Teixeira et al., 1995). Resistance can develop rapidly after the introduction of a new antimicrobial, with resistance to Linezolid and Daptomycin having both been reported in *S. aureus* strains within 2 years of introduction (Mangili et al., 2005; Tsiodras et al., 2001) Also of concern is the emergence of MRSA with reduced susceptibility for vancomycin (Chang et al., 2003; Hiramatsu et al., 1997), a glycopeptide typically reserved as a last line of treatment, although prevalence of these resistant strains has so far remained low (Fridkin et al., 2003).

1.2.3 Community-associated MRSA (CA-MRSA)

Since the 1990s, MRSA that is capable of causing severe disease in young otherwise healthy individuals, rather than immunocompromised hospitalised patients has emerged in the community, and is often found in those in close contact with others such as athletes, children, and military personnel (Begier et al., 2004; Herold et al., 1998; Morrison-Rodriguez et al., 2010). This often manifests as community-onset skin and soft tissue infections (Herold et al., 1998; King, 2006), although more severe disease associated with CA-MRSA also includes necrotising pneumonia (Francis et al., 2005; Gonzalez et al., 2005; Hidron et al., 2009) and necrotising fasciitis (Miller et al., 2005).

Since the emergence of CA-MRSA over the past two decades, dissemination has occurred worldwide and CA-MRSA is now an endemic problem in many countries, including the UK and USA (Dukic et al., 2013; Mediavilla et al., 2012; Otter and French, 2008; Tenover and Goering, 2009). In the United States, the USA300 clone is the dominant epidemic CA-MRSA clone (Tenover and Goering, 2009), and multiple CA-MRSA clones have been reported in many countries in Europe (Otter

and French, 2010). CA-MRSA appear to be largely distinct from the hospital-associated strains, prompting questions as to their evolutionary origins. However, in recent years, evidence for clonal spread and crossover of strains between community and hospital settings has emerged (David et al., 2014; Kourbatova et al., 2005; Saiman et al., 2003; Song et al., 2011).

1.3 *S. aureus* in animals

1.3.1 Host range and clinical significance

S. aureus colonises a wide range of animal species including domestic livestock such as cows, sheep, goats, pigs and poultry (Aarestrup, Cavaco, and Hasman, 2010; Hasman et al., 2010; Huijsdens et al., 2006; Rabello et al., 2007; Smyth et al., 2009), and companion animals, including dogs, cats, rabbits and horses (Vancraeynest et al., 2004; Walther et al., 2008; Weese et al., 2005). In addition, *S. aureus* has been isolated from wildlife including mammals such as wild boar, wild dolphins, wild birds and non-human primates (Morris et al., 2011; Porrero et al., 2013; Schaumburg et al., 2012).

As with humans, *S. aureus* infection can result in a range of clinical manifestations in these different species, and in particular in livestock it can cause disease of huge welfare and economic importance (McNamee et al., 1998; Miles, Lesser, and Sears, 1992). For example, *S. aureus* is one of the major causative agents for bovine contagious mastitis, and is spread between cows in the milking parlour affecting one or more quarters of the mammary gland, which is often reflected by elevated somatic cell counts in bulk milk tank samples (Barkema et al., 1998; Unnerstad et al., 2009; Nickerson, Owens, and Boddie, 1995; Waage et al., 1999). Disease can be chronic, subclinical and endemic within the herd, with low cure rates observed using typical intramammary antimicrobial preparations (Owens et al., 1997; Sol et al., 2000).

Pathogenesis of *S. aureus* in mastitis involves initiation of infection through invasion of the teat canal, followed by adherence to and invasion of bovine mammary epithelial cells and multiplication within host tissue (Almeida et al., 1996; Sutra and Poutrel, 1994). Failure of antimicrobial treatment and low cure rates may be attributed to factors such as intracellular persistence and biofilm formation (Cucarella et al., 2004; Hébert et al., 2000). Not only does this have welfare implications but it is economically important for the dairy industry, with costs associated with culling, treatment and lost milk yields estimated to be £300 million per year in the UK alone (Hillerton and Berry, 2005). Understanding the source and pathogenesis of such pathogens is important in order to design specific therapeutic interventions and effective control measures.

1.3.2 Livestock-associated MRSA (LA-MRSA)

Since the first isolation of MRSA from a case of bovine mastitis in 1972 (Devriese, Van Damme and Fameree, 1972), MRSA has been isolated from a wide range of animal species, including horses, pigs, poultry and cattle (Anderson et al., 2009; Hasman et al., 2010; Meemken et al., 2010). Livestock-associated MRSA (LA-MRSA), predominantly a clone known as CC398, is associated primarily with pigs, poultry and veal calves (Battisti et al., 2010; Graveland et al., 2010; Monecke et al., 2013). It is widely distributed in Europe, with studies showing the prevalence of CC398 LA-MRSA on pig farms to be as high as 39 % in a survey of 540 slaughter pigs in the Netherlands (de Neeling et al., 2007) and 56 % in a study of 50 pig farms (Van Den Broek et al., 2009). Other countries have also shown a high prevalence in pigs, with one study in Canada finding nearly 25 % of 285 pigs on 20 pig farms positive for CC398 LA-MRSA (Khanna et al., 2008).

Importantly, LA-MRSA is also of public health concern due to transmission to people in regular close contact with livestock, with the first report in 2003 in a family

and co-workers living on a pig farm, in which all isolates in pigs and humans were found to be identical by genotyping methods (Huijsdens et al., 2006). Workers such as pig farmers and abattoir personnel have been shown to be at an increased risk of nasal colonisation with LA-MRSA, with nasal carriage rates often significantly higher than those seen in the general population (Graveland et al., 2010; Van Den Broek et al., 2009; Voss et al., 2005). For example, a study on a Dutch poultry farm and slaughterhouse found that a higher percentage of personnel were colonised (5.6 %) in comparison to the general Dutch population (0.1 %) and that of these, personnel in direct contact with live animals were at increased risk of exposure (Mulders et al., 2010). In recent years, observations were made of people without animal contact acquiring LA-MRSA (van Loo et al., 2007; Wulf et al., 2008). CC398 LA-MRSA has since been suggested to have originated from a human-adapted population of CC398 MSSA, after which the livestock strains acquired methicillin and tetracycline resistance (Price et al., 2012).

Until recently, all MRSA contained the gene for methicillin resistance, known as *mecA*, on a mobile genetic element (MGE) known as SCC*mec* (Utsui and Yokota, 1985; Ito et al., 1999) (See section 1.5.2.3). However, since 2011 *S. aureus* strains containing a divergent methicillin resistance gene known as *mecC* (previously designated $mecA_{LGA251}$) have been described (García-Álvarez et al., 2011; Shore et al., 2011). Surveys have identified *mecC*-positive MRSA in 13 European countries such as the UK, Ireland, Denmark, Germany and France (Cuny et al., 2011; García-Álvarez et al., 2011; Laurent et al., 2012; Shore et al., 2011). MRSA containing *mecC* is predominately identified in ruminant-associated clones CC130 and CC425, however it has also been isolated from human infections, and has been found in at least 14 different host species to date (García-Álvarez et al., 2011; Gómez et al., 2014; Loncaric et al., 2013; Paterson et al., 2012; Porrero et al., 2014; Walther et al., 2012). For example, the bovine-associated CC130 clone containing *mecC* has been isolated primarily from ruminants and humans, but in addition, has also been isolated from wildlife including a chaffinch and seal, and domestic animals including a cat, dog and guinea pig (Medhus et al., 2013; Paterson et al., 2012; Walther et al., 2012;

Wieler et al., 2011). (For further information see section 1.5.2.3, section 1.6.1 and Table 1.1).

In addition, there are sporadic reports of *S. aureus* strains that are usually associated with humans being isolated from cases of bovine mastitis, indicating that the transmission dynamics of *S. aureus* can move in both directions (Lowder et al., 2009; Sakwinska et al., 2011). Therefore, these findings highlight the need to understand the evolutionary origins and transmission dynamics of strains of *S. aureus*, including MRSA which colonise multiple hosts.

1.4 *S. aureus* population genetics and evolution

1.4.1 *S. aureus* molecular typing methods

Biotyping methods and genetic molecular typing have been used to study *S. aureus* isolated from animals and humans and have been highly informative for determining the structure of natural *S. aureus* populations (Devriese, 1984; Enright et al., 2000; Musser and Selander, 1990; Smyth et al., 2009). Prior to the development of genotyping methods, phenotypic biotyping was used to differentiate "ecovars" from diverse host species such as humans, ruminants and poultry (Devriese, 1984). For example, the ruminant strains could be distinguished from those of humans or poultry based on the ability to coagulate ruminant plasma (Devriese, 1984). One of the earliest studies to find out about natural *S. aureus* populations in different hosts was a large-scale study employing Multi-Locus Enzyme Electrophoresis (MLEE) on over 2000 *S. aureus* isolates isolated from bovine and ovine mastitis isolates, and human MRSA (Musser and Selander, 1990). This study identified 14 major *S. aureus* electrophoretic types, with the majority of strains in each type associated with a single host species (Musser and Selander, 1990). Other molecular typing methods of *S. aureus* include pulsed field gel electrophoresis (PFGE) (Murchan et al., 2003), multi-locus variable number of tandem repeat analysis (MLVA) (Francois et al.,

2005; Sabat et al., 2003), and *spA* typing, which assesses the allelic variation and number of repeats in the polymorphic region of the Staphylococcal protein A (*spa*) gene (Shopsin et al., 1999).

A widely used typing technique is Multi-locus Sequence Typing (MLST) (Enright et al., 2000). This method utilises polymorphic variation observed in the DNA sequence of regions within seven housekeeping loci to assign numbers to each allelic variant observed, the combination of which denotes the sequence type (ST) (Enright et al., 2000). Closely related STs differing at only 1 or 2 loci are termed single (SLV) and double locus variants (DLV) respectively, and can be grouped together in clonal complexes (CC) using the eBURST algorithm, with each CC referred to generally as a "clone" (Feil et al., 2004). In addition to *S. aureus*, MLST typing schemes have been designed for a wide range of pathogenic bacteria (http://pubmlst.org/databases/). The *S. aureus* MLST database currently houses over 2800 different STs (http://saureus.mlst.net/, last accessed on 28th February 2014). MLST has been widely used to study the population genetics of *S. aureus* including epidemiologically important human MRSA and MSSA clones (Enright et al., 2002), *S. aureus* from animal hosts including livestock and wildlife (Porrero et al., 2013; Smyth et al., 2009) and to compare animal and human isolates to identify potential transmission events (Armand-Lefevre, Ruimy and Andremont, 2005; Sakwinska et al., 2011).

MLST has been complemented in some studies by the use of *sas* typing, which also assesses allelic variation in regions of 7 genes, the combination of which denotes a *sas* type (Robinson and Enright, 2003). However, the *sas* genes (designated *sasA, sasB, sasD, sasE, sasF, sasH,* and *sasI*) encode *S. aureus* surface-associated proteins containing the LPXTG motif. For example, *sasA* encodes an adhesin (SraP) that mediates platelet binding, and has been shown to affect virulence *in vivo* in a rabbit endocarditis model (Siboo, Chambers and Sullam, 2005). As such these genes are predicted to be more rapidly evolving with elevated heterogeneity, thereby providing additional resolution, although analysis revealed that dN/dS ratios indicated no evidence for diversifying selection (Robinson and Enright, 2003). However, *sas*

typing has been used successfully to examine host specificity in *S. aureus* isolates from a range of animal species, and was found to largely delineate the strains according to host (Smyth et al., 2009).

1.4.2 *S. aureus* population structure

Molecular typing studies employing the methods described above have determined that the natural population structure of *S. aureus* is highly clonal, in contrast to other bacterial species in which recombination is much more frequent than mutation (Feil et al., 2003; Spratt, Hanage and Feil, 2001). Clonal complexes have been named according to the ancestral genotypes predicted by the BURST and eBURST algorithms (Feil et al., 2003, 2004). In evolutionary studies clones are often referred to in the context of their evolutionary lineages and may be referred to as a "clonal lineage" (Cuny et al., 2010; McAdam et al., 2012; Stegger et al., 2013).

Many clonal lineages are largely specific for a single host species with the majority of isolates represented by only a few predominant lineages (Fitzgerald et al., 1997; Musser et al., 1990; Reinoso et al., 2008; Zadoks et al., 2000) For example, Musser *et al*, found that a single MLEE electrophoretic type was responsible for over 88 % of urogenital toxic shock syndrome cases in humans (Musser et al., 1990). The major lineages associated with HA-MRSA are relatively few in number, and have been identified as CC5, CC8, CC22, CC30, CC45 and CC239 (Enright et al., 2002; Ko et al., 2005; Lindsay et al., 2006; Xu et al., 2009). In humans, replacement of predominant clones over time in the hospital setting has been observed. In a study that traced MRSA genotypes within a German hospital over an 11 year period, 2 major MRSA clones known as CC5/ST228-MRSA-I and CC45-MRSA-IV (Berlin EMRSA) were almost completely displaced by epidemic clones CC22-MRSA-IV and CC5-MRSA-II (Albrecht et al., 2011). The major CA-MRSA lineages are described as CC1, CC8, CC30, CC59, CC80 and CC93 (Tristan et al., 2007; Vandenesch et al., 2003). Certain lineages can predominate in specific geographical regions, for example, CC80 CA-MRSA is commonly found in many European

countries including Norway (Francis et al., 2005), Germany (Monecke et al., 2006), the Netherlands (Stam-Bolink et al., 2007), and Sweden (Fang et al., 2008), whereas in the United States, the USA300 (ST8) clone is widely distributed (Moran et al., 2006).

The major animal-associated lineages have been identified to include ST5, ST71, ST97, ST126, ST130, ST133, ST151, ST425, ST771 and ST873 (Smyth et al., 2009; Sung, Lloyd and Lindsay, 2008). Of these, the major ruminant-adapted genotypes have been described as CC97, CC151, ST425 CC130, and CC133 (Guinane et al., 2010; Smith et al., 2005; Smyth et al., 2009; Sung, Lloyd and Lindsay, 2008).

While the majority of *S. aureus* genotypes are largely restricted to a single host, including ruminant-adapted genotypes such as CC151 and CC133, there are some genotypes that appear to have a wider host tropism. In addition to CC398 described in section 1.3.2, which colonises pigs, veal calves, poultry and humans, there are ruminant-associated genotypes such as CC97 and CC130 which have also been isolated from humans (Ellington et al., 2008; García-Álvarez et al., 2011; Harrison et al., 2013; Smith et al., 2005), and the widely disseminated ST5 genotype is found in both humans and poultry (Lowder et al., 2009).

Phylogenetic analysis comparing ruminant-associated genotypes with major human genotypes indicate that the ruminant clones occupy distinct and narrowly distributed clades within the phylogenetic tree, indicating that they have human ancestors, and that there may have been several host jumps into ruminants (Figure 1.1) (Fitzgerald, 2012; Guinane et al., 2010). However, there are some recent phylogenetic studies using MLST data indicating possible host jumps back into humans (Shepheard et al., 2013; Weinert et al., 2012), highlighting the need to determine the evolutionary origins of clones with a wide host range.

High throughput sequencing of multiple bacterial genomes has expanded on molecular typing techniques by enabling phylogenetic analysis based on core genome single nucleotide polymorphisms (SNPs) (Comas et al., 2013; Harris et al.,

2010; Harrison et al., 2014; He et al., 2010; McAdam et al., 2012). These methods have provided higher resolution, facilitating the evolutionary analysis of subtypes within a single MLST sequence type (Harris et al., 2010). For example, clonal expansion of the HA-MRSA ST239 clone has been observed on a global scale, with the population structure clustering into geographic subclades corresponding to Asia and South America, with more diverse basally situated European isolates indicating a possible European origin (Harris et al., 2010). There was also evidence of several intercontinental transmission events (Harris et al., 2010). An expanded dataset resolved ST239 further to the level of country and city, providing evidence that ST239 isolates from Turkey arose from a single introduction from eastern Europe (Castillo-Ramírez et al., 2012).

Transmission dynamics of MRSA between hospitals can be identified, such as the observed frequent transmission of ST239 MRSA between Turkish hospitals in Ankara and Istanbul, whereas a single founder event followed by limited transmission resulted in a distinct clade of ST239 in an Izmir hospital, which was consistent with metadata concerning patterns of human movement (Castillo-Ramírez et al., 2012). The human-associated CC30 lineage was resolved into at least 3 major clades that represent the major pandemic clones phage type 80/81, southwest pacific (SWP) and EMRSA-16, with phylogenetic analysis based on core genome SNPs indicating that each clone evolved independently, rather than sequentially, as suggested by previous studies based on MLST data (McAdam et al., 2012; Robinson et al., 2005). Even within the CC30 EMRSA-16 clone, Bayesian phylogenetic analysis of UK isolates resolved several subclades specific for particular regions of the UK, with hospitals in major cities acting as a reservoir for transmission to other regions (McAdam et al., 2012).

Retrospective genome sequencing of outbreaks of infectious diseases has great potential for clinical diagnostic and disease surveillance measures within hospitals (Köser et al., 2012). For example, using phylogenetic analysis of core genome SNPS, Köser *et al*, were able to clearly differentiate between strains of MRSA involved in an outbreak in a neonatal unit, and other MRSA strains within the same hospital that

were not involved in the outbreak, in addition to identifying resistance and virulence gene profiles (Köser et al., 2012). In the same hospital in 2011, a persistent outbreak of MRSA was attributed to a novel sequence type (ST2371), revealing transmission within the neonatal unit, between mothers, and into relatives in the community (Harris et al., 2013). Rapid benchtop sequencing facilitated real-time identification of MRSA outbreak strains despite an extended 64-day period of no MRSA strains being isolated, and identified re-infection as the result of a persistently colonised staff member (Harris et al., 2013).

The evolution of *S. aureus* strains within a single host can also be evaluated (McAdam et al., 2011; Young and Golubchik, 2012). For example, the genomes of 3 sequential *S. aureus* strains isolated over a 26 month period from a cystic fibrosis patient indicated the existence of a heterogeneous population of bacteria that had evolved from a single infecting strain (McAdam et al., 2011). Young *et al* used high throughput sequencing to identify the genetic changes accompanying the transition from a nasal carriage ST15 MSSA isolate to a fatal bacteraemia infection in a single individual, in which an excess of protein-truncating mutations were identified as preceding disease progression (Young and Golubchik, 2012).

The mutation rate for several lineages of *S. aureus* based on core genome SNPS have been estimated as 3.3×10^{-6} substitutions per site per year for ST239 (Harris et al., 2010), 1.42×10^{-6} for CC30 (McAdam et al., 2012), 2.2×10^{-6} for ST225 (Nübel et al., 2010) and 5.125×10^{-6} for CC5 (Lowder et al., 2009). Using these rates of molecular evolution has enabled the dates for the emergence of new clones to be estimated (Harris et al., 2010; McAdam et al., 2012). The date for the most recent common ancestor of the ST239 clade is estimated to be in the mid to late 1960s, contemporaneous with the emergence of methicillin resistance (Harris et al., 2010; Jevons, 1961). The CC30 lineage is predicted to have emerged in 1842, with emergence of the phage type 80/81, southwest pacific and EMRSA-16 clones estimated to be in 1936, 1967 and 1975 respectively (McAdam et al., 2012). In the case of the 80/81 CC30 subclade, this preceded the widespread therapeutic use of penicillin, indicating that acquisition of resistance may have been a possible selective

reason for successful expansion of this clone (McAdam et al., 2012). The timing for host jumps that *S. aureus* has undergone has also been estimated (Lowder et al., 2009) (For more detail see section 1.6.2). Overall, phylogenetic analysis of *S. aureus* using core genome SNPs has proven to be a successful method to resolve in more detail the evolution of certain lineages of *S. aureus*.

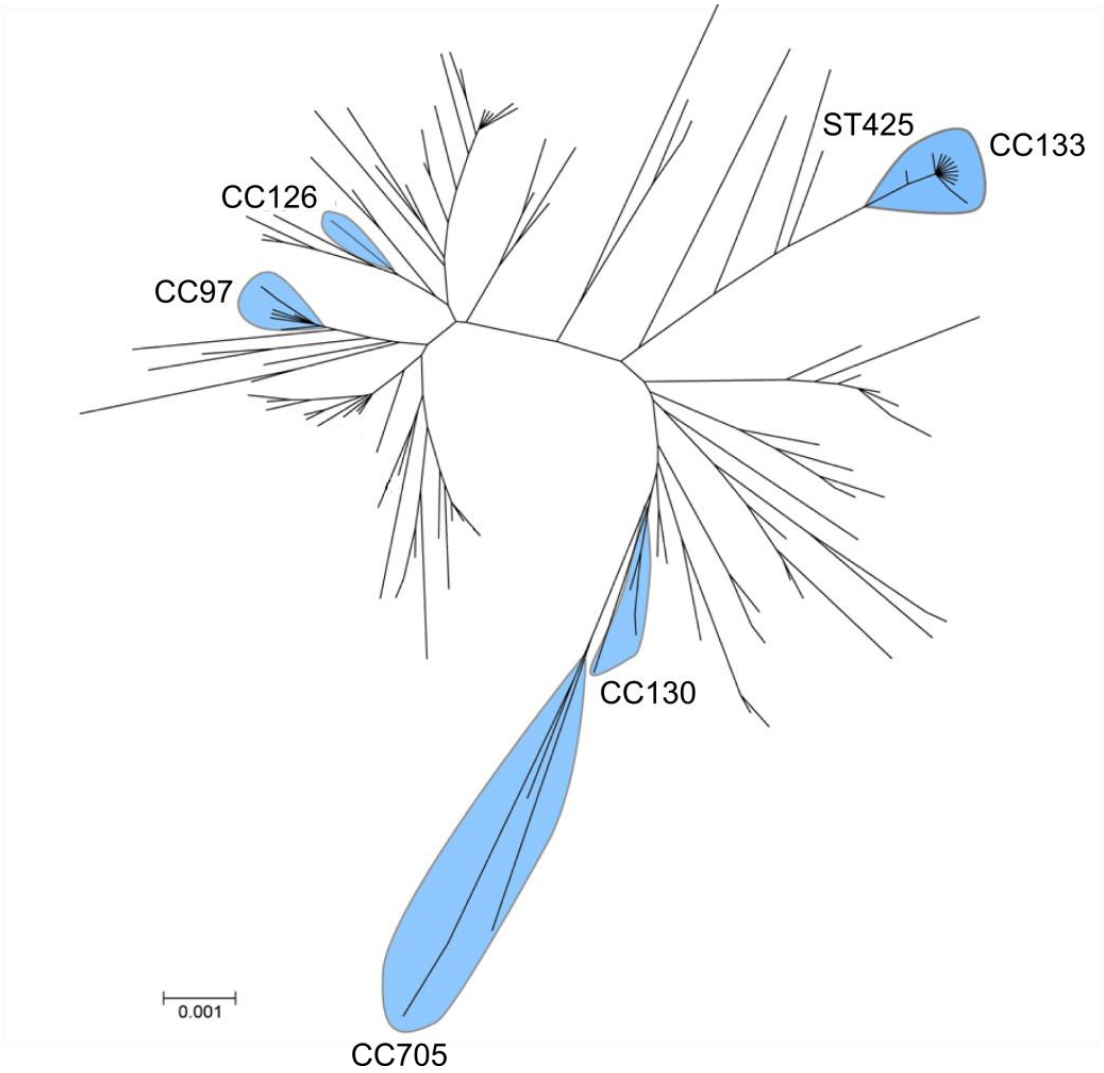


Figure 1.1. Neighbour joining tree for *S. aureus* constructed using concatenated MLST sequence data. Included in the tree are representative common animal and human-associated genotypes. The major ruminant clades are highlighted in blue with clonal complexes labelled as shown. Adapted from (Fitzgerald, 2012).

1.4.3 Role of recombination in *S. aureus* evolution

Homologous recombination involving the horizontal exchange of genetic material between different strains of bacteria is a method by which strains can generate genetic diversity, and such recombination events can range in size from a short DNA sequence (Wilson et al., 2011) through to large segments of the *S. aureus* genome that are several hundred thousand kilobases in length (Robinson and Enright, 2004). The relative contributions that recombination and mutation make towards the diversification and evolution of different bacterial species varies (Feil et al., 1999; Feil, Enright and Spratt, 2000; Feil et al., 2003). As previously stated, *S. aureus* is highly clonal, with estimates based on MLST data indicating that genetic divergence of alleles is 15 times more likely to occur by point mutation than by recombination (Feil et al., 2003). In contrast, other bacterial species are highly recombinant, such as *Neisseria meningitidis* and *Streptococcus pneumoniae*, in which recombination is predicted to contribute to allelic variation 4 times and 10 times more frequently than point mutation in each bacterial species respectively (Feil et al., 2000).

However, recent recombination analysis of a range of *S. aureus* clonal complexes indicates that recombination in the core genome may be more widespread than previously identified, and, despite the clonal population structure of *S. aureus*, recombination is observed to occur to varying degrees in the accessory genome, core and core variable genome (Castillo-Ramírez et al., 2011; Everitt et al., 2014; Lindsay et al., 2006; Watanabe et al., 2009).

The widespread horizontal transfer of MGE among *S. aureus* has resulted not only in strain variation in terms of the presence or absence of MGE, but also variation in genetic content within MGE as a result of homologous recombination (Lindsay et al., 2006). Homologous recombination among different SCC*mec* elements has resulted in mosaic structures among different SCC*mec* types both in terms of allelic variation and gene content variation (Ito et al., 2001) (For additional information on SCC*mec* see section 1.5.2.3). Extensive genetic diversity also exists among *S. aureus* phages

indicated by their mosaic gene contents that are likely to be as a result of recombination events (Kwan et al., 2005).

Core variable genes as described by Lindsay *et al*, often encode surface-expressed and secreted proteins known to interact with host and as such are subject to host selective pressures, which may promote diversification of the gene by homologous recombination (Lindsay et al., 2006). Recombination has been observed for several core variable adhesion genes between different clonal complexes (Basic-Hammer et al., 2010). For example, staphylocoagulase is an extracellular virulence factor of *S. aureus* that coagulates plasma by mediating binding to prothrombin, and examination of multiple staphylocoagulase (*coa*) gene sequences indicates evidence of recombination resulting in extensive allelic variation (Watanabe et al., 2009). The genes *fnbA* and *fnbB* encode fibronectin binding proteins FnBPA and FnBPB that promote *S. aureus* colonisation through interaction with host ligands such as fibrinogen, elastin and fibronectin (Burke et al., 2010; Peacock et al., 2000). Examination of the 7 allelic variants of the A domain of FnBPB indicates that the same isotypes can be found in phylogenetically distinct strains, indicative of horizontal transfer by homologous recombination (Burke et al., 2010). Examination of gene sequences encoding FnBPA from different lineages also indicates that recombination has taken place, with evidence of mosaic gene structure in the *fnbpA* gene (McCarthy and Lindsay, 2010).

A recently characterised novel core genome-encoded superantigen, SElX, which is widely distributed in *S. aureus* also demonstrates considerable allelic diversity consistent with genetic reassortment as a result of homologous recombination (Wilson et al., 2011). SElX bypasses normal antigen presentation to cause non-specific mitogenicity through binding MHC Class II molecules outside the normal binding groove on antigen presenting cells and in addition, binding to the variable region of the T cell receptor β-chain (Wilson et al., 2011). The recombination that has resulted in the allelic variation of *selx* may contribute to the overlapping but

distinct Vβ activation profiles observed in human and bovine lymphocytes, in which varying Vβ subgroups are activated in response to SElX (Wilson et al., 2011).

Congruence analysis of gene trees for MLST genes of *S. aureus*, which are core genome-encoded housekeeping genes and predicted to be under neutral selective pressure, indicates that recombination has occurred over the long term to an extent that it can disrupt phylogenetic signal (Feil et al., 2001). Recombination involving the core genome of *S. aureus* has been observed to play a fundamental role in the emergence of some clonal complexes (Robinson and Enright, 2004). In particular, the widespread and successful HA-MRSA CC239 clone has a mosaic genome containing an import of approximately 500 kbp to 600 kbp from a CC30 donor that has transferred horizontally into a CC8 genetic background, in a region of the genome that spans the origin of replication (Holden et al., 2009; Robinson and Enright, 2004; Smyth et al., 2010). Of note, within the CC239 clone, the rate of recombination in MGE and core genome has been observed to vary between different phylogeographic subclones (Castillo-Ramírez et al., 2012). Large-scale homologous recombination events of approximately 250 kbp in size have also been observed to create hybrid genomes for ST34 and ST42, with both imports associated with the region of the genome spanning the origin of replication (Robinson and Enright, 2004; Thomas et al., 2012). The mechanism for these large-scale chromosomal recombination events has not yet been elucidated (Robinson and Enright, 2004). Given the large size of the sequences involved, it has been hypothesised that conjugation may be the most likely mechanism (Robinson and Enright, 2004). The incidental transfer of regions of core genome that flank MGE has also been proposed as a driver for certain core genome recombination hotspots, including the region spanning the origin of replication (Everitt et al., 2014). Homologous recombination enables bacteria to rapidly drive genetic divergence as an adaptive response to selective pressures imposed by different ecological niches, and as such, hybrid genomes involving large recombinational replacements have also been observed in other clinically relevant bacterial species such as *Acinetobacter baumanii* (Snitkin et al., 2011), *Escherichia coli* (Touchon et al., 2009),

Streptococcus agalactiae (Brochet et al., 2008), and *Streptococcus pneumoniae* (Hiller et al., 2010).

One mechanism that controls the horizontal import of DNA in *S. aureus* for both acquisition of MGE and recombination involves the restriction modification systems (RM) of *S. aureus*. In *S. aureus,* the type 1 RM system consists of a single restriction gene (*hsdR*) which is located distantly to 2 copies each of modification (*hsdM*) and specificity (*hsdS*) genes found in genomic islands νSaα and νSaβ (Kuroda et al., 2001; Roberts et al., 2013; Waldron and Lindsay, 2006). Combinations of these enzymes form subunits that methylate host DNA at specific target recognition sequences which then protect it from the endonuclease which recognises and cleaves unmethylated DNA containing the same target recognition sequence (Tock and Dryden, 2005; Waldron and Lindsay, 2006). *S. aureus* strains have conserved copies of *hsdM* and *hsdR* genes, but the *hsdS* gene that specifies the target recognition sequence has been found to vary according to clonal complex, such that each lineage recognises 2 specific target recognition sequences (Waldron and Lindsay, 2006). This is due to the mosaic structure of the target recognition domains in proteins encoded by *hsdS*, which is likely to have arisen due to multiple duplication and recombination events (Roberts et al., 2013; Waldron and Lindsay, 2006). To date the target recognition sequences for the major human clones CC1, CC5, CC8 and ST239 have been determined (Roberts et al., 2013). It has been proposed that the type 1 RM system of *S. aureus* has resulted in the observed lineage-specific distribution and content of MGE and core variable genes in different clonal complexes (McCarthy and Lindsay, 2010, 2012; Waldron and Lindsay, 2006). In particular, assessment of the distribution of target recognition sequences in large conjugative plasmids that carry multiple antimicrobial resistance genes indicated that certain plasmids have evolved to reduce the number of type 1 RM target recognition sites in order to facilitate exchange in particular from CC8 to CC5 lineages (Roberts et al., 2013). More recently a type IV RM system (SauUSI) has been described, which recognises cytosine methylated DNA and protects against exogenous uptake of DNA (Corvaglia et al., 2010; Monk et al., 2012; Xu et al., 2011). *S. aureus* is not naturally competent, and in particular, ST97 has been shown to be refractory to transformation using *in*

vitro laboratory manipulations (Monk et al., 2012). However, under SigH staphylococcal sigma factor-activating conditions, natural competence of *S. aureus* has been demonstrated *in vitro*, indicating that horizontal acquisition of exogenous DNA is possible (Morikawa et al., 2012). To date, the complement of RM genes in ST97 has not yet been fully characterised.

Overall, studies investigating homologous recombination in *S. aureus* have concentrated on the major human-associated *S. aureus* lineages, and it is therefore not yet known to what extent recombination events such as the large-scale regions identified in CC239, contributes towards the evolution of animal-associated *S. aureus* lineages.

1.5 Comparative genomics of *S. aureus*

1.5.1 Sequencing technologies

The first published complete bacterial genome sequence was of a *Haemophilus influenzae* strain, published in 1995, and completed using conventional Sanger sequencing (Fleischmann et al., 1995). Since 2005, the development of several different high throughput massively parallel sequencing technologies have been developed, bringing the cost of bacterial genome sequencing down and increasing the speed and practicality of sequencing multiple bacterial genomes quickly (Metzker, 2010).

High throughput sequencing technologies vary according to platform, which can differ according to library preparation, amplification stage and sequencing chemistry. For example, Illumina (http://www.illumina.com/) uses a cyclic reversible terminator chemistry in which fluorescently-labelled reversible terminator nucleotides emitting different wavelengths are perfused across the flow cell, any complementary labelled nucleotides binding to the template are imaged, then cleaved for the next cycle (Metzker, 2010). Roche 454 uses pyrosequencing technology, in which cycles of a

single type of dNTP are washed across the template, and each time a nucleotide that is complementary to the template is incorporated by DNA polymerase, the release of pyrophosphate due to an enzyme cascade is measured as visible light, the intensity of which indicates the number of nucleotides that are being incorporated against the template (Metzker, 2010). Real time sequencing from a single molecule has been developed by Pacific Bioscience (http://www.pacificbiosciences.com/) in which real-time imaging of fluorescently labelled nucleotides occurs as they are incorporated by a DNA polymerase along the length of a single template molecule (Eid et al., 2009). Resulting sequences can range in read length from just over 30 base pairs through to hundreds or even thousands of base pairs (Metzker, 2010).

The first complete *S. aureus* genomes, sequenced using traditional Sanger sequencing, were MRSA strains N315 and Mu50 (which was also resistant to vancomycin) published in 2001 (Kuroda et al., 2001). There are now 40 complete genomes and over 2600 draft *S. aureus* genomes in the Genbank database (http://www.ncbi.nlm.nih.gov/genbank/), and over 14000 *S. aureus* raw genome sequence reads available in the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra), with this number expected to expand hugely in coming years. Genomes that have been sequenced are often those of clinical interest in human medicine, and as such, of the 40 complete *S. aureus* genomes, only 5 are from animals, including the ruminant strains RF122 (NC_007622), ED133 (NC_017337), LGA251(NC_017349), avian strain ED98 (NC_013450) and CC398 LA-MRSA S0385, although isolated from a human endocarditis case (García-Álvarez et al., 2011; Guinane et al., 2010; Herron-Olson et al., 2007; Lowder et al., 2009; Schijffelen et al., 2010). Further genome sequencing of animal strains are required if the genetics underlying the multiple host associations of *S. aureus* are to be elucidated.

High throughput sequencing has many applications in *S. aureus* research including population genetics and phylogenetic studies (Castillo-Ramírez et al., 2012; Harris et al., 2010). Recently, benchtop sequencing platforms such as the Ion Torrent (http://www.lifetechnologies.com), Illumina Miseq (http://www.illumina.com) and

454 GS Junior (http://454.com) have been developed for rapid sequencing, with the intended aim for use in real-time epidemiological surveillance studies of pathogen outbreaks and in clinical diagnostics (Köser et al., 2012). For example, bench-top sequencers including Ion Torrent were used in a crowd-sourced open access collaborative effort to trace the origins of an outbreak of the shiga-toxin producing *E. coli* O104:H4 (STEC) that caused a serious outbreak of haemolytic-uraemic syndrome in Europe in 2011 (Mellmann et al., 2011; Rohde et al., 2011).

Overall, high throughput sequencing has provided important information regarding the genome organisation, population genetics and evolution of important *S. aureus* clones.

1.5.2 Genome organisation of *S. aureus*

Genome sequences have provided important information regarding the genome structure and evolution of *S. aureus*. Comparison of *S. aureus* genomes from different lineages and host species has enabled the identification of genetic features that may relate to the successful emergence of certain clones, or adaptation of those strains to a particular ecological niche.

S. aureus genomes are circular chromosomes that range in size from 2692570 bp to 3043219 bp and are on average 2828269 bp in length, with some strains containing additional extrachromosomal plasmids (http://www.ncbi.nlm.nih.gov/genome/). The genomes are low in GC content, approximately 33 %, and contain approximately 2700 genes (Baba et al., 2002; Fitzgerald et al., 2001; Holden et al., 2004; Kuroda et al., 2001). Comparative genomic hybridisation studies and comparison of *S. aureus* genomes show that there is a core "backbone" of the genome comprising groups of orthologous genes that are shared among *S. aureus*, ranging from 70 % to 78 % (Figure 1.2) (Zakour et al., 2008; Fitzgerald et al., 2001; Lindsay and Holden, 2004). Approximately 10-12 % of the genome is comprised of "core-variable" genes, which are variably present among strains (Lindsay et al., 2006). The combinations of core

variable genes are lineage-specific, and many encode surface-expressed or secreted proteins that may interact with host, for example the genes involved in capsule production and the gene encoding collagen adhesin (Lindsay et al., 2006). In a DNA microarray study that examined regions of difference between 12 human, bovine, ovine and caprine *S. aureus* strains, 180 CDS were found to be host-specific for ruminants, and of these over 50 % encoded predicted secreted or surface-expressed proteins, that may be involved in the interaction of host and pathogen (Zakour et al., 2008).

The greatest genetic variation observed among strains is in the complements of "accessory genome" that they contain, primarily comprising MGE, which undergo frequent exchange between strains through horizontal gene transfer (HGT) (Figure 1.2). The major MGE contained in *S. aureus* are plasmid, Staphylococcal pathogenicity islands (SaPI), bacteriophage, and Staphylococcal cassette chromosomes (SCC) which are summarised below with regard to structure, content, and contribution towards virulence or host adaptation (Sections 1.5.2.1-5).

The distribution and variation in genetic content of *S. aureus* MGE can be marked, even within strains of the same lineage. Many clinically-important *S. aureus* clones have shown variation in MGE content, such as CC5, ST22, and CC30 (Holden et al., 2013; Lowder et al., 2009; McAdam et al., 2012). For example, the distribution of MGEs within strains of the CC398 lineage varied according to host (pig or human) and the country in which they were isolated (Belgium, Denmark or the Netherlands) (McCarthy et al., 2011). There have also been examples of multiple acquisition events of different SCC*mec* types within the same lineage including ST5 (Nübel et al., 2008), CC8 (Robinson and Enright, 2003) and CC22 (Boakes et al., 2011).

1.5.2.1 Bacteriophages

S. aureus phages are widely distributed, with most strains carrying at least one phage (Baba et al., 2002; Brüssow, Canchaya and Hardt, 2004; Canchaya et al., 2003; Goerke et al., 2009). In fact, traditionally, bacteriophages were used in a method for

typing bacterial strains, by assessing the patterns of lysis in strains with a panel of different phages (Wentworth, 1963). The majority of *S. aureus* phages belong to the *Siphoviridae* family which are approximately 40 kbp in size, and others belong to the *Podoviridae* (class I, less than 20 kbp in size) and *Myoviridae* (class III, over 125 kbp) families (Kwan et al., 2005; Vybiral et al., 2003). The *Siphoviridae* family of phages can be further classified according to integrase gene sequence and insertion sites (Goerke et al., 2009; Lindsay and Holden, 2004). In addition to the virulence genes listed below, staphylococcal bacteriophages have conserved functional modules associated with lysogeny, DNA metabolism, DNA packaging and capsid morphogenesis, tail morphogenesis and host cell lysis, although recombination has resulted in extensive mosaic content between different phages (Kwan et al., 2005).

Phages encode several known virulence factors in *S. aureus,* including bi-component cytotoxin Panton-Valentine leukocidin (PVL), encoded by the genes *lukF-PV, lukS-PV,* and widely associated with CA-MRSA (Vandenesch et al., 2003), as well as bovine leukocidin genes *lukM, lukF* which encode lukM/lukF'-PV, that has been shown to have the greatest cytotoxic activity against bovine neutrophils (Barrio, Rainard and Prévost, 2006). The gene encoding Exfoliative toxin A (ETA), which causes scalded skin syndrome is also encoded on a phage (Yamaguchi et al., 2000). Genes including staphylokinase (*sak*), staphylococcal complement inhibitor (*scn*), staphylococcal chemotaxis inhibitory protein (*chp*) and staphylococcal enterotoxin A (*sea*), collectively known as the immune evasion complex (IEC) are contained in various combinations on the β-toxin converting phage, and known to be important in evading human innate immune defenses (for more details see section 1.7.1) (Wamel et al., 2006).

Frequent gain and loss of bacteriophages contributes towards genome plasticity and adaptation of *S. aureus* strains to changing environments, and can lead to subpopulations of strains within a single infection demonstrating different virulence properties (Goerke et al., 2006; Goerke and Wolz, 2004).

1.5.2.2 Staphylococcal pathogenicity islands (SaPI)

Staphylococcal pathogenicity islands (SaPI) are a family of chromosomal islands 14-17 kbp in length that have several well-defined features to describe them, including that they can be grouped according to specific attachment (*att*) sites and sequence homology of integrase genes (*int*), are mobilised and horizontally transferred between bacterial strains by bacteriophages, and encode a variety of virulence genes including superantigens (Lindsay et al., 1998; Novick, 2003). The SaPI structure has conserved regions consisting of the site-specific integrase gene (*int*), transcriptional regulators, replication genes, a phage encapsidation module, and a variable region consisting of virulence genes including Toxic Shock Syndrome Toxin-1 (TSST-1, encoded by *tst*), Biofilm-associated protein (*bap*), staphylococcal enteroxins SEK, SEQ, SEC (*sek, seq, sec*), and an allelic variant of von Willebrand factor binding protein (*vwb*) (Figure 1.3) (Lindsay et al., 1998; Novick, Christie and Penadés, 2010; Novick, 2003; Úbeda et al., 2003; Viana et al., 2010). SaPIs have been found in *S. aureus* genomes from both human and animal strains (Figure 1.3), some of which contain genes encoding proteins that have been found to have host-specific functions (Section 1.7.1) (Guinane et al., 2010; Kuroda et al., 2001; Lindsay and Holden, 2004).

Excision of SaPIs occurs through the induction of temperate phages, which replicate and efficiently package the SaPI into small phage-like particles (Lindsay et al., 1998; Maiques et al., 2007). SaPIs can be packaged by more than one type of phage, and can be transferred at high frequency (Lindsay et al., 1998; Ruzin, Lindsay and Novick, 2001; Úbeda et al., 2005). The excision-replicating-packaging cycle has been shown to be initiated by the binding of a non-essential phage protein to the global repressor Stl which controls SaPI gene expression, thereby initiating the transfer mechanism (Tormo-Más et al., 2010).

1.5.2.3 Staphylococcal cassette chromosome (SCC)

The most widely distributed SCC element is SCC*mec* containing the *mecA* gene which encodes an altered penicillin binding protein (PBP2a) with a reduced affinity for binding β-lactam antimicrobials, thereby allowing bacteria to resist bacteriocidal killing by continued cross-linking of bacterial cell wall peptidoglycans (Utsui and Yokota, 1985). SCC*mec* elements integrate at a specific insertion site located at the 3' end of the *orfX* gene which encodes a rRNA large subunit methyltransferase (Boundy et al., 2013). SCC*mec* gene organisation and content is highly variable but all contain a '*mec* gene complex' and '*ccr* gene complex', domains which form the basis for SCC*mec* typing as outlined by the International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC, 2009). The *mec* complex comprises *mecA,* and regulatory proteins *mecR1* encoding a signal transducer protein and *mecI* encoding a repressor protein (IWG-SCC, 2009). Variation in this complex is due to insertion sequences IS431 or IS1272 disrupting the genes or regions that are upstream and downstream, and has resulted in the description of class A, B, and C *mec* gene complexes (Katayama, Ito and Hiramatsu, 2001). The *ccr* gene complex contains the *ccr* genes, which contain considerable sequence variation and are classified into 3 main types including *ccrA, ccrB* (both of which contain subtypes) and *ccrC,* which are responsible for the site and orientation-specific excision and integration of SCC*mec* (Ito et al., 2004; Katayama, Ito and Hiramatsu, 2000). SCC*mec* elements can be subtyped according to joining regions that reside upstream, downstream and inbetween the *mec* and *ccr* gene complexes, and in some cases these regions carry additional resistance or virulence determinants such as genes encoding heavy metal and antibiotic resistance (Li et al., 2011).

To date, 11 different SCC*mec* types have been described, designated SCC*mec*-I to SCC*mec*-XI (Berglund et al., 2008; García-Álvarez et al., 2011; Ito et al., 2001; Ito et al., 2004; Li et al., 2011; Ma et al., 2002; Oliveira, Milheiriço and de Lencastre, 2006; Shore, et al., 2011; Zhang et al., 2009). SCC*mec* types I-III are often associated with HA-MRSA and contain additional antimicrobial resistance genes on integrated plasmids and transposons situated in the joining regions (Hiramatsu et al.,

2001; Holden et al., 2004; Ito, Katayama and Hiramatsu, 1999). For example, SCC*mec* type II has an integrated copy of plasmid pUB110 containing resistance genes for tobramycin, kanamycin (*aad*) and bleomycin (*ble*), and transposon Tn554 containing resistance genes for erythromycin (*ermA*) and spectinomycin (*spe*) (Matthew T. G. Holden et al., 2004; T Ito et al., 1999). SCC*mec* types IV and V are commonly described in CA-MRSA isolates (Daum et al., 2002; Ito et al., 2004; Okuma et al., 2002; Vandenesch et al., 2003).

SCC*mec* types IX, X and V(5C2and5)c have been characterised from CC398 LA-MRSA (Li et al., 2011). SCC*mec* type V(5C2and5)c contains the *czr* gene, conferring resistance to cadmium and zinc, and is widespread among CC398 LA-MRSA isolated from animals and humans, with the use of zinc in animal feed proposed as a reason for the co-selection of methicillin resistance in this lineage (Cavaco et al., 2010; Cavaco, Hasman and Aarestrup, 2011; Price et al., 2012).

The most recently discovered SCC*mec* type is SCC*mec* type XI, which contains a divergent *mecA* homolog that has 63 % amino acid identity when compared to *mecA* from SCC*mec* type III (García-Álvarez et al., 2011; Shore, et al., 2011). This novel *mecA* gene, now termed *mecC*, is of public health concern, as MRSA containing *mecC* can be erroneously designated MSSA according to standard *mecA*-specific or PBP2a-specific PCR and monoclonal antibody tests and may therefore by under-reported (García-Álvarez et al., 2011) (See section 1.3.2 for additional information).

SCC elements which contain the *ccr* gene complex but not the *mec* gene complex have been described in *S. aureus*, some of which contain virulence or resistance genes. For example, human strain MSSA476 has an SCC element containing a fusidic acid resistance gene (Holden et al., 2004). SCC elements containing a capsule type 1 operon (SCC*cap1*) and resistance genes for mercury (SCC*Hg*) have also been described (Chongtrakool et al., 2006; Luong et al., 2002).

The arginine catabolic mobile element (ACME) was first described in the USA300 lineage (ST8) associated with SCC*mec*-IVa and containing genes for an arginine

deiminase pathway (*arc* operon) and oligopeptide permease system (opp-3) (Diep et al., 2006). It is widely distributed within the USA300 lineage (Diep et al., 2008). Variants of ACME that lack the opp-3 gene cluster have also been found in small numbers in other human-associated lineages including ST5 (USA100-SCC*mec*-II), ST59 (USA1000-IVa), ST22 (SCC*mec*-IV) and ST97-SCC*mec*-V (Diep et al., 2008; Ellington et al., 2008; Shore et al., 2011b).

A study utilising ACME deletion mutant strains in a rabbit bacteraemia infection model indicated that ACME does contribute towards virulence (Diep et al., 2008), although a separate study indicated that ACME has no effect in rodent pneumonia and skin and soft tissue infection models (Montgomery, Boyle-Vavra and Daum, 2009). A more recent study indicates that the arginine deiminase system encoded by the *arc* operon on ACME promotes growth and survival in acidic conditions that mimic that of human skin (Thurlow et al., 2013). In addition, ACME contains SpeG, a polyamine acetyltransferase which promotes persistence of infection by resisting host polyamine production in the anti-inflammatory wound healing stage of skin infections (Thurlow et al., 2013). To date, ACME has not been described in *S. aureus* strains isolated from animals. In general, this region of the genome has been shown to have plasticity among *S. aureus* strains, and can involve the acquisition of a variety of elements including type I restriction modification systems and enterotoxin genes (Everitt et al., 2014; Noto et al., 2008).

1.5.2.4 Plasmids

S. aureus may contain one or more plasmids which are circularised auto-replicating DNA molecules that may reside independently in the cell, or linearise and integrate into the host genome, such as in human strain MRSA252 (Holden et al., 2004). A recent study on over 250 *S. aureus* strains revealed that 90 % carried at least one plasmid, and of these, 75 % were 20-30 kbp in size (Shearer et al., 2011). Three main groups of plasmids have been described ranging in size from 1.3 kbp to 65 kbp, and varying in replication mechanism (Malachowa and DeLeo, 2010; Shearer et al., 2011). Class I small multicopy plasmids are typically less than 5 kbp in size, undergo

rolling circle replication and often carry a single resistance gene determinant (Khan, 2005).

Class II larger low copy plasmids, range in size from 15 kbp to 30 kbp and encode several resistance determinants on integrated plasmid and transposons, and are likely to be horizontally transferred by transduction (Lindsay and Holden, 2006). A common plasmid-encoded resistance gene is the *blaZ* gene specific for beta-lactamase which hydrolyses the beta-lactam ring of penicillin, and, in addition to regulatory genes *blaI* and *blaR*, may be encoded on plasmids, the chromosome, transposons and SCC*mec* (Olsen, Christensen and Aarestrup, 2006; Shore et al., 2011). Other resistance genes include those for antiseptics (*qac*) and heavy metals such as cadmium, arsenic and mercury (Baba et al., 2002; Kuroda et al., 2001). Genes encoding toxins such as exfoliative toxin B and bacteriocin immunity protein have also been found on plasmids, such as pRW001 (Jackson and Iandolo, 1986) in addition to enterotoxin genes responsible for food-borne illness (Shearer et al., 2011). Class II plasmids have recently been subdivided into 3 subgroups based on comparison of plasmid sequences from multiple strains including pIB485-like enterotoxin plasmids, pMW2-like plasmids found in animal and human strains, and pUSA300HOUMR-like plasmids, also found in human and animal strains (Shearer et al., 2011).

Class III are large conjugative plasmids such as pSK41, which are over 40 kbp in size and contain multi-resistance genes for antimicrobials, antiseptics and disinfectants, in addition to the *tra* genes that enable conjugative transfer (Berg et al., 1998; Lindsay and Holden, 2006). For example, vancomycin-resistant *S. aureus* (VRSA) acquired a conjugative plasmid from Enterococci which contains transposon 1546 encoding the *vanA* operon (Weigel, 2003).

1.5.2.5 Genomic islands

Genomic islands vSaα and vSaβ appear to be present in the great majority of *S. aureus* strains but contain extensive variation in gene content (Lindsay and Holden,

2006). Both islands contain putative transposases and type 1 restriction-modification genes including *hsdM* and *hsdS* (Baba et al., 2002; Kuroda et al., 2001; Waldron and Lindsay, 2006). Genomic island α contains superantigen-like (*ssl*) genes and genes encoding a lipoprotein (*lpl*) gene cluster, while genomic island β contains the bi-component leukotoxin genes *lukD* and *lukE,* a serine protease gene cluster, in addition to genes encoding hyaluronate lyase, superantigen toxins and bacteriocin (Lindsay and Holden, 2004). There is variation in the gene content and organisation of the genomic islands among *S. aureus* strains from different lineages (Lindsay et al., 2006; Lindsay and Holden, 2004).

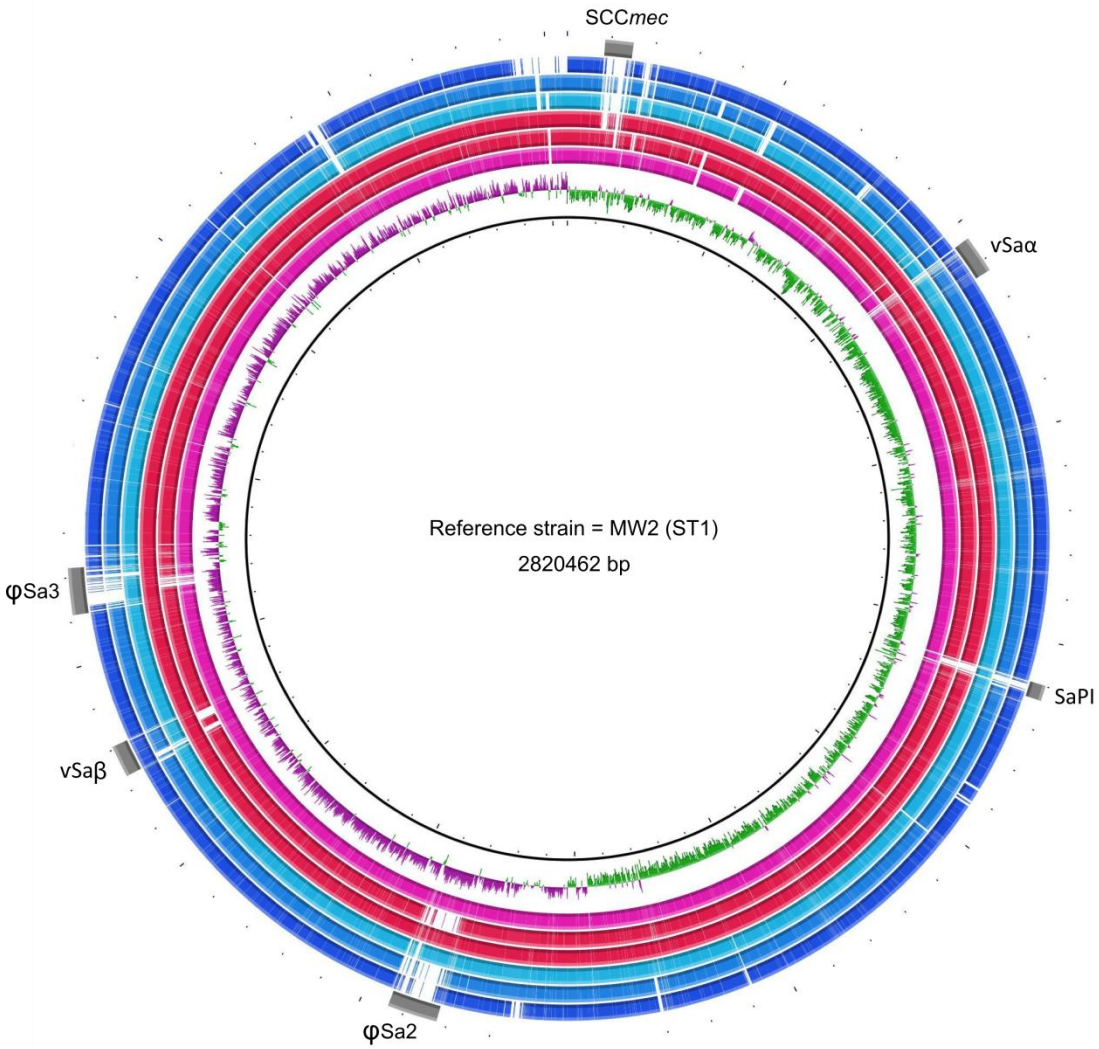


Figure 1.2: Genome organisation in *S. aureus.* Circular schematic diagram based upon pairwise BLASTN sequence comparison of 3 human (red rings) and 3 ruminant (blue rings) *S. aureus* strains. To demonstrate the conserved core "backbone" of *S. aureus*, genomes are aligned against a human ST1 reference strain, MW2 (NC_003923). The rings are labelled as follows extending from the inner ring to the outer ring: 1) GC skew, 2) USA300_TCH1516 (human ST8), 3=N315 (human ST5), 4=MSSA476 (human ST1), 5= ED133 (ovine ST133), 6=LGA251 (bovine ST425), 7=RF122 (bovine ST151), and 8= MW2 MGE illustrated as examples.

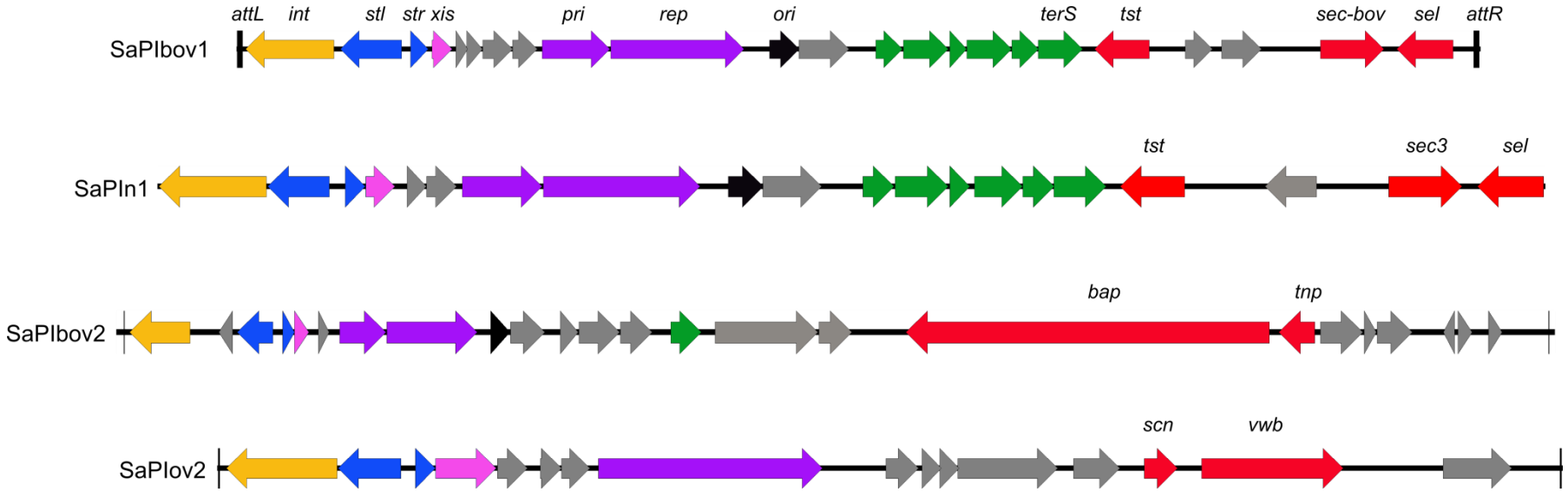


Figure 1.3. Schematic diagram illustrating a selection of *S. aureus* SaPIs from different host species. SaPIbov1 and SaPIbov2 are from bovine *S. aureus* strains (Fitzgerald et al., 2001; Úbeda et al., 2003), SaPIov2 from ovine *S. aureus* strain ED133 (Guinane et al., 2010) and SaPIn1 from human ST5 strain N315 (Kuroda et al., 2001). Left and right attachment sites are shown (*attL* and *attR*). Genes are colour-coded and labelled at the top as follows: genes for hypothetical proteins are grey, integrase gene (*int*) is yellow, transcriptional regulators (*stl*, *str*) are dark blue, excisionase (*xis*) is pink, replication module genes (*pri*, *rep*) are purple, packaging genes including the gene for terminase small subunit (*terS*) are green. Virulence genes are highlighted red, including genes for TSST-1 (*tst*), bovine and human allelic variants of enterotoxin c (*sec-bov* and *sec3*), staphylococcal enterotoxin l (*sel*), biofilm associated protein (*bap*), transposase (*tnp*), staphylococcal complement inhibitor (*scn*) and von Willebrand factor binding protein (*vwb*).

1.6 *S. aureus* multiple host colonisation

Although *S. aureus* clones are largely adapted to a specific host species, there are a number of clones that appear to display a wider host tropism (Table 1.1). Molecular studies indicate that that there are several methods by which a strain is capable of colonising multiple host species, including zoonotic and anthroponotic transmission events, in addition to strains that are capable of jumping hosts and adapting to an entirely different host environment, resulting in divergent host-adapted clones (Harrison et al., 2013; Lowder et al., 2009; Van Duijkeren et al., 2004). This is also supported by phenotypic data indicating variation in infectivity of strains from different hosts (Moodley et al., 2012). However, it is currently unknown as to why some clones remain specific while others appear to be more "generalist" in terms of host colonisation.

Table 1.1: *S. aureus* lineages associated with multiple host species

ST	Humans	Cattle	Small ruminants	Pigs	Poultry	Horses	Companion animals*	Wildlife	References
ST1	✓	✓		✓	✓	✓		✓	(Franco et al., 2011; Hasman et al., 2010; Lindsay et al., 2006; Lowder et al., 2009; Schaumburg et al., 2012; Smyth et al., 2009; Sung, Lloyd and Lindsay, 2008)
CC5	✓	✓	✓	✓	✓	✓	✓		(Couto et al., 2012; Hata et al., 2010; Khanna et al., 2008; Lowder et al., 2009; Sasaki et al., 2012; Smith et al, 2013)
CC8	✓	✓		✓		✓	✓	✓	(Haenni et al., 2012; Himsworth et al., 2014; Lin et al., 2011; Lindsay et al., 2006; Sakwinska et al., 2011; Walther et al., 2009)

ST	Humans	Cattle	Small ruminants	Pigs	Poultry	Horses	Companion animals*	Wildlife	References
ST9	✓		✓	✓	✓			✓	(Porrero et al., 2012; Fessler et al., 2011; Hasman et al., 2010; Mulders et al., 2010; Neela et al., 2009; Schaumburg et al., 2012)
ST22 (EMRSA-15)	✓		✓			✓	✓		(Porrero et al., 2012; Harrison et al., 2014; Moodley et al., 2006)
CC97	✓	✓	✓	✓				✓	(Battisti et al., 2010; Ellington et al., 2008; Gharsa et al., 2012; Himsworth et al., 2014; Rabello et al., 2007)
ST121	✓						✓(Rabbits)		(Lozano et al., 2011; Smyth et al., 2009)

ST	Humans	Cattle	Small ruminants	Pigs	Poultry	Horses	Companion animals*	Wildlife	References
CC130	✓	✓	✓					✓	(Porrero et al., 2012; Cuny et al., 2011; García-Álvarez et al., 2011; Gómez et al., 2014)
ST133		✓	✓				✓ (Cats)	✓ (wild boar)	(Porrero et al., 2012; Guinane et al., 2010; Hasman et al., 2010; Meemken et al., 2013; Sasaki et al., 2012; Smyth et al., 2009)
ST398	✓	✓	✓	✓	✓	✓		✓	(Abdelbary et al., 2014; Fessler et al., 2010; Hasman et al., 2010; Himsworth et al., 2014; Price et al., 2012; Smith et al., 2013; Uhlemann et al., 2012)

* Companion animals refers to cats and dogs specifically in this table, unless stated otherwise.

1.6.1 *S. aureus* zoonotic transmission through contact between animals and humans

The CC398 LA-MRSA clone is capable of colonising a wide range of host species including pigs, turkey, veal calves, cattle, small ruminants and horses (Table 1.1). Transmission can occur between livestock and people in regular contact with them, with pig farmers, slaughterhouse workers and veterinarians shown to be at an increased risk of colonisation with LA-MRSA CC398 (Lefevre et al., 2005; Garcia-Graells et al., 2012; Van Cleef et al., 2010; Voss et al., 2005). The colonisation has been linked to the duration of livestock contact (Graveland et al., 2011) and this clone does not appear to be easily transmissible between humans, since lower carriage rates are observed for in-contact household members of pig farmers (Garcia-Graells et al., 2013). There is evidence that human-adapted subclones exist, since CC398 has been isolated from humans with no prior livestock contact, with phylogenetic evidence for a human MSSA ancestral state (McCarthy et al., 2012; Price et al., 2012; Uhlemann et al., 2012) (Section 1.6.2). There is public health concern about zoonotic transmission of other clones such as the pig-associated ST9 clone (Lefevre et al., 2005; Neela et al., 2009), and a genome sequencing study identified 2 cases of on-farm zoonotic transmission of the *mecC*-positive MRSA CC130 clone (Harrison et al., 2013). Taken together, these findings indicate that livestock are an important reservoir for the transmission of *S. aureus* to humans that are in regular contact with them.

In addition, predominantly human-associated genotypes such as CC1, CC5 and CC8 are isolated from cases of bovine mastitis, indicating that transmission may occur in both directions between livestock and humans (Aires-de-Sousa et al., 2007; Hata et al., 2010; Rabello et al., 2007). For example, although the direction could not be firmly established, MRSA ST1 strains of spa type t127 that had been previously identified in humans were found in cows and a farmer in close contact, which were

indistinguishable by phenotypic and genotypic tests (Juhász-Kaszanyitzky et al., 2007).

Horses have been found to be colonised with MRSA, with outbreaks often reported with post-operative cases in veterinary hospitals (Van Duijkeren et al., 2010; Weese et al., 2005). Reports of transmission of *S. aureus* between horses and humans in close contact with them has also been reported (Van Duijkeren et al., 2010; Weese et al., 2005b, 2006). The CC8 clone and in particular the Canadian MRSA-5 clone (ST8) is the predominant clone identified in horses (Moodley et al., 2006; Weese et al., 2005b; Weese et al., 2006c; Walther et al., 2009). The CC8 subtypes isolated from horses are typically distinct from the CC8 human strains isolated in the same geographic region, suggesting adaptation to the horse environment rather than episodes of reverse zoonoses (Cuny et al., 2006; Moodley et al., 2006; Weese et al., 2006c). In addition, a CC398 subtype has been identified that is associated with horses (Abdelbary et al., 2014) (For more details see Section 1.6.2).

Although *Staphylococcus pseudintermedius* is the primary Staphylococcal species associated with dogs, companion animals including dogs and cats may be colonised with *S. aureus* including MRSA (Hanselman et al., 2009; Loeffler et al., 2011). Transmission of strains that are indistinguishable by molecular typing methods appears to occur in both directions between companion animals and humans, although the genotypes involved often reflect the predominant human-associated genotype in that region (Malik et al., 2006; Moodley et al., 2006; O'Mahony et al., 2005; Weese et al., 2006). For example, in the USA, the predominant clones identified in companion animals are the human-associated clones USA100 (CC5) and USA300 (CC1) (Lin et al., 2011). A predominant human CA-MRSA clone, known as epidemic MRSA-2 has been isolated from cats and dogs in suspected transmission events with in-contact humans (Weese et al., 2006). In the UK, dogs have been found to be colonised with the human epidemic HA-MRSA clone EMRSA-15 (CC22) (Baptiste et al., 2005; Loeffler et al., 2005; Moodley et al.,

2006). A recent phylogenetic genomic study comparing ST22 human and ST22 UK companion animal isolates found that the companion animal isolates were interspersed throughout the EMRSA-15 pandemic clade and clustered with the UK human EMRSA-15, indicating frequent exchange of isolates between humans and companion animals from the same MRSA population (Harrison et al., 2014). Companion animals may be acting as a reservoir for humans and vice versa, with transmission observed within the veterinary setting similar to that observed in human hospitals (Harrison et al., 2014). Human isolates occupied a basal position on the phylogenetic tree, indicating that the evolutionary origin of the companion animal strains was from humans (Harrison et al., 2014) There were no SNPs significantly associated with host adaptation, suggesting that the EMRSA-15 ST22 clone is an "extended spectrum genotype" capable of colonising multiple hosts (Harrison et al., 2014).

1.6.2 Host shifts followed by adaptation to a new host

Previous genome sequencing and phylogenetic studies have identified certain *S. aureus* lineages that appear to have undergone a host jump during the past, followed by adaptation to the new host, resulting in distinct host-adapted sublineages (Lowder et al., 2009; Price et al., 2012). For example, the avian-associated ST5 lineage was found to have arisen as the result of a single host jump event from a likely human polish ancestral ST5 strain, followed by adaptation to the avian host and global dissemination via the poultry industry, with Bayesian phylogenetic analysis indicating this occurred approximately 40 years ago (Lowder et al., 2009). Data from a phylogenetic study conducted by Price *et al*, in which the evolutionary history of CC398 was reconstructed, indicated that a human-associated MSSA subclone occupied a basal position on the phylogenetic tree, suggesting that a human-to-livestock host jump followed by acquisition of methicillin resistance resulted in the emergence of livestock-associated CC398 MRSA (Price et al., 2012). Further, a recent phylogenetic study based upon SNP discovery at 97 housekeeping gene loci

indicates that there may be a horse-associated sublineage of CC398 circulating in Europe (Abdelbary et al., 2014). However, further genomic data would be required to determine if this subclade has undergone a host jump and adapted to the equine host, since this subclade also contained strains from other species including humans who had been in contact with horses, which may indicate transmission events (Abdelbary et al., 2014).

Evidence for more ancient host jumps from humans into animals suggests that humans have been an important source for the emergence of *S. aureus* lineages that have become endemic in animal hosts (Guinane et al., 2010; Weinert et al., 2012). For example, major ruminant-associated lineages CC151 and CC133 appear to have emerged from a human-associated host a long time ago and have since evolved to become ruminant-specific lineages (Guinane et al., 2010; Weinert et al., 2012). Bayesian phylogenetic analysis using MLST sequence data from over 100 strains from different hosts indicated that multiple host jumps from humans to ruminants have occurred, leading to the emergence of the major ruminant-associated clades CC133, CC151 and CC97, with the earliest switch estimated to have occurred approximately 5500 years ago, consistent with the cattle domestication and expansion of agricultural practices in the Old World (Weinert et al., 2012). The human-to-ruminant host jump resulting in the emergence of CC133 is estimated to have occurred approximately 3000 years ago, with the other major ruminant-associated clones of CC97 and CC151 estimated to have emerged approximately 1800 and 5400 years ago, respectively (Weinert et al., 2012). In a large-scale host-association study of over 3000 isolates using MLST data, 13 anthroponoses were identified, which, in addition to the human-to-bovine and human-to-poultry host jumps already described above, also identified a human-to-rabbit host switch (Shepheard et al., 2013).

Both studies also indicate that 2 clones circulating in human populations, including CC59 and CC25 may have arisen from animal-associated ancestors (Shepheard et al.,

2013; Weinert et al., 2012). CC59 is a major human-associated clone in Taiwan and Japan (Song et al., 2011; Wang et al., 2012), and is estimated to have arisen at least 60 years ago from a ruminant-associated ancestor (Weinert et al., 2012). While studies employing MLST sequence data are very useful for long-term evolutionary studies, the sequence data is not detailed enough to identify very recent host jumps, and as such, phylogenetic studies employing high throughput sequencing have greater resolution (Price et al., 2012). It is important to determine more recent host switch events in order to identify newly emerging clones of public and veterinary health significance.

1.7 Molecular basis for host adaptation

1.7.1 Horizontal transfer of MGE

Comparative genomic studies have identified several determinants encoded on MGE that are associated with host specificity in human and animal hosts, including proteins involved with antimicrobial resistance, virulence or evasion of host immune defenses (Table 1.2) (Guinane et al., 2010; Herron-Olsen et al., 2007; Lowder et al., 2009; Wamel et al., 2006). Comparative genomic hybridisation studies comparing animal and human *S. aureus* have found that the core genome is highly conserved despite occupying diverse hosts, and that many of the core variable genes identified in human strains are also present in animal strains, suggesting that variation in MGE content may play a central role in host specificity (Zakour et al., 2008; Sung, Lloyd and Lindsay, 2008).

For example, MGEs appear to have contributed to ruminant host adaptation (Guinane et al., 2010; Herron-Olson et al., 2007). In bovine CC151 strain RF122, MGEs including phages and SaPI were identified that had not previously been observed in human strains (Herron-Olson et al., 2007). Similarly in ovine CC133 strain ED133,

novel MGEs including 2 SaPI and 3 phages were identified (Guinane et al., 2010). This includes a SaPI containing the *vwb* gene, which is a variant of the chromosomal gene encoding von Willebrand factor binding protein (vWbp) (Figure 1.3) (Guinane et al., 2010; Viana et al., 2010). Variants of this SaPI have been identified in ruminant *S. aureus* strains of several different genotypes, including CC97, CC133 and CC398, and have been demonstrated to confer the ability to specifically coagulate ruminant plasma (Guinane et al., 2010; Viana et al., 2010). An equine SaPI containing an allelic variant of vWbp was identified in a single ST133 equine strain and found to coagulate equine and ruminant plasma, suggesting a possible role in either host (Viana et al., 2010).

SaPIbov2 carries biofilm associated protein Bap (encoded by the *bap* gene) which mediates binding and intercellular adhesion in biofilm formation (Cucarella et al., 2001), and although it is less widely distributed than the biofilm-forming *ica* operon, Bap is thought to contribute to persistence in the mammary gland (Cucarella et al., 2004; Úbeda et al., 2003) (Figure 1.3). The gene encoding toxic shock syndrome toxin TSST-1 (*tst*) and superantigen enterotoxin genes are encoded on SaPIs in human *S. aureus* strains and in ruminant strains (Figure 1.3). Host-specific allelic variants of TSST-1, and enterotoxin genes *sec* and *sel* are encoded on SaPIbov1 and SaPIov1 (Fitzgerald et al., 2001; Guinane et al., 2010), while human strains contain distinct SaPIs that contain *tst* along with a variety of superantigens including allelic variants of *sec*, *sel*, *seb*, *seq*, *sek* and a pathogenicity island protein known as *ear* (Baba et al., 2002; Lindsay et al., 1998; Lindsay and Holden, 2004).

A bacteriophage has been identified in bovine strains that encodes a bi-component leukotoxin known as lukM/lukF'-PV (Herron-Olson et al., 2007), which has been shown to have elevated cytotoxic activity for bovine neutrophils (Barrio et al., 2006). In a study examining the prevalence of *S. aureus* leukotoxins in animal *S. aureus* isolates from Japan, the prevalence of the genes for lukM/lukF'-PV were found to be 62.5-86.1 % in bovine mastitis isolates and 0 % in avian and porcine isolates

(Yamada et al., 2005). The prevalence was lower in a smaller study of ruminant isolates in Spain, in which 35.4 % of bovine isolates encoded lukM/lukF'-PV, and 53.1 % in bovine isolates from Germany and Switzerland, suggesting regional variation in carriage of this MGE (Monecke et al., 2007; Rainard et al., 2003). The highest prevalence of this phage has been observed in small ruminant isolates from sheep and goats isolated in France at 96.9 % and 68.1 % respectively (Rainard et al., 2003).

Some livestock-associated strains have been found to contain novel antimicrobial resistance genes contained on multi-resistance plasmids, which likely reflects the selection pressures imposed by antimicrobial use, particularly in the pig industry (Fessler, Kadlec and Schwarz, 2011; Kadlec and Schwarz, 2009). For example, a novel apramycin resistance gene, *apmA*, which is contained on a plasmid discovered in bovine, porcine and a single poultry ST398 LA-MRSA, has not as yet been identified in human *S. aureus* (Fessler, Kadlec and Schwarz, 2011; Fessler et al., 2011). The novel *vga(C)* gene has also been discovered in LA-MRSA CC398 of porcine origin in a multi-resistance plasmid, and confers resistance to streptogramin A, lincosamides and pleuromutilins (Kadlec and Schwarz, 2009). This gene has also been described in bovine and poultry CC398 isolates (Fessler et al., 2010; Fessler et al., 2011).

Following the human-to-poultry host jump of CC5 *S. aureus* approximately 40 years ago, the emergent clone underwent host adaptive evolution, which included the acquisition of MGE from a novel avian-specific accessory genome pool, leading to the globally dominant CC5 clone currently found in poultry (Lowder et al., 2009). This included the acquisition of 2 novel prophages, 2 plasmids and a SaPI which contained avian-specific genes of unknown function, suggestive of a central role for MGE in adaptation to the avian host (Lowder et al., 2009). A novel phage of the β-converting phage family known as φAvβ, was widely distributed in avian strains of several genotypes, including CC5, CC385 and CC398 (Lowder et al., 2009; Price et

al., 2012). This phage lacked any of the genes that are identified on human β-converting phages, which are collectively termed the Immune Evasion Cluster (IEC), and are known to encode proteins that are important in evading human innate immune defenses (Wamel et al., 2006). The human IEC genes include *sak, scn, chp, sea* and *sep* encoding staphylokinase, staphylococcal complement inhibitor, chemotaxis inhibitory protein and variants of staphylococcal enteroxins type A and type P respectively (For further detail see below). Instead, the φAvβ phage contained an ornithine cyclodeaminase and putative novel protease (Lowder et al., 2009). An avian-specific plasmid, pAvX was widely distributed among the poultry CC5 strains, which, in addition to a lysophospholipase gene, contained the *scpA* gene which encodes a cysteine (thiol) protease that has been previously identified in a chicken with dermatitis and found to be avian-specific (Lowder et al., 2009; Takeuchi et al., 2002).

Perhaps one of the most widely studied MGE relating to host specificity is the human-associated β-converting phage, also known as ϕSa3 (Goerke et al., 2009; Verkaik et al., 2010; Wamel et al., 2006). Lysogenic conversion occurs in which the phage integrates into the β-toxin gene, inactivating it, however, it brings into the strain the IEC genes (as listed above) (Wamel et al., 2006). Staphylokinase (*sak*) is anti-opsonic and decreases neutrophil phagocytosis due to the degradation of human opsonins IgG and C3b on the bacterial surface (Rooijakkers et al., 2005). Staphylococcal complement inhibitor (*scn*) acts on C3 convertases to inhibit all pathways of the complement system and has been shown to have human-specific activity (Rooijakkers et al., 2005). Chemotaxis inhibitory protein (*chp*) impairs leukocyte chemotaxis and migration through interfering with formylated peptide and C5b receptors (Haas et al., 2004; Postma et al., 2004). Staphylococcal enteroxin A (*sea,* and a variant *sep*) is also commonly encoded on the β-toxin converting phage (Wamel et al., 2006). Staphylococcal enterotoxin A is a superantigen that causes non-specific activation of T lymphocytes, however it also down-modulates binding of chemokines to the cell surface of monocytes through proteolytic cleavage of chemokine receptors, thereby inhibiting chemotaxis (Rahimpour et al., 1999).

Distinct combinations of IEC genes have been found in different *S. aureus* phages with 7 variants described so far (Wamel et al., 2006). Given the fitness advantage the presence of this phage would likely confer in human hosts, it is not surprising that over 90 % of human *S. aureus* surveyed contain this phage (Verkaik et al., 2010; Wamel et al., 2006). In a study by Goerke *et al*, investigating phage dynamics in infection (cystic fibrosis patients and bacteraemia cases) and nasal carriers, 96 % of nasal *S. aureus* had a stable β-converting phage, while in infection isolates extensive phage dynamics were observed including translocation to distant genomic sites, duplication and recombination events (Goerke et al., 2006). In contrast 82 % of bovine strains have an intact β-toxin gene, and lack the β-converting phage (Monecke et al., 2007).

Table 1.2. Genes associated with specific host species contained on *S. aureus* MGE.

Gene	Protein	MGE	Function (if known)	Host specificity	Reference
vwb	Von Willebrand factor binding protein (vWbp)	SaPI (SaPIov2, SaPIeq1)	Coagulation of ruminant and equine plasma	Ruminant, equine	(Guinane et al., 2010; Viana et al., 2010)
bap	Biofilm-associated protein (Bap)	SaPIbov2	Biofilm formation	Ruminant	(Cucarella et al., 2001, 2004; Úbeda et al., 2003)
lukM/lukF	Bi-component leukotoxin lukM/lukF'-PV	Bacteriophage	Greatest leukotoxic activity on bovine neutrophils	Ruminant	(Barrio et al., 2006; Herron-Olson et al., 2007; Rainard et al., 2003)
apmA	Aminocyclitol acetyltransferase	Plasmid	Apramycin resistance	Livestock-associated MRSA CC398 (bovine, porcine, poultry)	(Fessler, Kadlec and Schwarz, 2011; Fessler et al., 2011)

Gene	Protein	MGE	Function (if known)	Host specificity	Reference
vga(C)	ABC transporter	Plasmid	Streptogramin A-lincosamide-pleuromutilin resistance	Livestock-associated MRSA CC398 (bovine, porcine, poultry)	(Fessler et al., 2011; Fessler et al., 2010; Kadlec and Schwarz, 2009)
*Lacks human IEC genes found in human β-converting phages	Ornithine cyclodeaminase and putative novel protease	φAvβ (β-converting phage family)	Unknown, possible avian-specific niche activity	Avian-poultry	(Lowder et al., 2009; Price et al., 2012)
scpA	Cysteine (thiol) protease	pAvX plasmid	Possible avian-specific niche activity	CC5 poultry	(Lowder et al., 2009; Takeuchi et al., 2002)
sak	Staphylokinase	β-converting phage	Anti-opsonin	Human	(Rooijakkers et al., 2005)
scn	Staphylococcal complement inhibitor (SCIN)	β-converting phage	Inhibits all pathways of the complement system	Human	(Rooijakkers et al., 2005)

Gene	Protein	MGE	Function (if known)	Host specificity	Reference
chp	Chemotaxis inhibitory protein (CHIPS)	β-converting phage	Interferes with formylated peptide and C5b receptors to inhibit leukocyte chemotaxis	Human	(Haas et al., 2004; Postma et al., 2004)
sea, sep	Staphylococcal enterotoxin A, (and variant staphylococcal enterotoxin P)	β-converting phage	Superantigens, and interferes with chemokine receptors	Human	(Rahimpour et al., 1999)

* The avian variant of the β-converting phage also inserts into the β-toxin gene, but lacks the immune evasion cluster (IEC) genes that are identified in the human variant of the β -converting phage, instead containing genes encoding proteins with unknown roles in avian pathogenesis, including ornithine cyclodeaminase and a putative novel protease.

1.7.2 Genome decay associated with adaptation to a new host

Adaptive genomic diversification in bacteria which have undergone a change to a new host or alteration in lifestyle such as transitioning to an intracellular lifestyle, may involve the loss of function of genes that are no longer beneficial in the new ecological niche (Moran and Plague, 2004). Gene function may be lost through a variety of mechanisms, including the acquisition of nonsense mutations that introduce a premature stop codon, or a small insertion or deletion event resulting in a frameshift mutation (Guinane et al., 2010; Herron-Olson et al., 2007; Lerat and Ochman, 2005). In some bacterial species, a gene may be inactivated through the insertion of a transposable element, a process which is utilised as a laboratory technique to create deletion mutants to assess the function of specific genes of *S. aureus in vitro* (Cucarella et al., 2001; Komatsuzawa et al., 2000; Lerat and Ochman, 2005). In several species of bacteria pseudogenes are primarily caused by frameshift mutations (Lerat and Ochman, 2005). However, in A+T rich bacteria such *as S. aureus*, a relatively higher proportion of pseudogenes are created by nonsense mutations compared to G+C rich bacteria, since stop codons are also A+T rich (Lerat and Ochman, 2005).

Pseudogenes in *S. aureus* contribute to a relatively small proportion of the genome, accounting for approximately 1.0 % to 1.5 % of genes, with the highest number of 88 pseudogenes reported in the ST36 human strain, MRSA252, and bovine CC151 strain RF122, which has 70 detected pseudogenes (Herron-Olson et al., 2007; Holden et al., 2004; Lerat and Ochman, 2005).

In comparison, the Gram positive bovine commensal and mastitis pathogen, *Streptococcus uberis* has 62 pseudogenes representing 3.4 % of the genome (Ward et al., 2009). *S. uberis* pseudogenes are largely represented by regulatory genes and genes for surface-expressed proteins, indicating possible continuing adaptation to the mammary gland (Ward et al., 2009). However, comparison of the *S. uberis* genome compared to other Streptococcal species indicates that the genome is well equipped

to acquire energy through different methods of carbohydrate metabolism, and contains polyphosphate kinase that may enable it to cope with the stress of transitioning between diverse ecological niches such as the rumen and the mammary gland (Ward et al., 2009).

Pseudogene formation is also characteristic of a transition from a free-living enteric bacterium such as *E. coli* to a human host-restricted intracellular pathogen such as *Shigella,* which has an abundance of pseudogenes accounting for up to 8 % of the genome as a result of substitutions and insertional inactivation by IS elements (Jin et al., 2002; Moran and Plague, 2004; Wei et al., 2003).

Comparison of the human leprosy pathogen *Mycobacterium leprae* to the related tubercle bacillus *Mycobacterium tuberculosis* indicates that since diverging from the last common ancestor, *Mycobacterium leprae* may have undergone extensive genome reduction, losing more than 2000 genes, and of what remains, over 50 % are pseudogenes, indicating that loss of gene function can be extensive in some pathogens (Cole et al., 2001).

Exploitation of an intracellular niche has been proposed as a mechanism of persistence in *S. aureus* mastitis pathogenesis (Garzoni and Kelley, 2009; Hébert et al., 2000). Genomic evidence for niche adaptation in ruminant *S. aureus* strains since they diverged from a human ancestor has been observed, including the loss of function of genes known to be important in human disease pathogenesis (Guinane et al., 2010; Harrison et al., 2013; Herron-Olson et al., 2007). In particular, of the 70 identified pseudogenes in CC151 bovine strain RF122, several included known human virulence genes, including staphylococcal protein A (*spa*) and clumping factor A (*clfA*), in addition to proteins involved in iron transport (*sstC*) and host interaction (*sdrC*) (Herron-Olson et al., 2007). The genome sequence of ovine CC133 strain, ED133 revealed pseudogenes in genes associated with metabolism, toxins, lipoproteins and within MGE (Guinane et al., 2010). Of these, 4 pseudogenes were found to be in common with bovine strain RF122, including genes for a lipoprotein, high affinity iron transporter, *splA* encoding serine protease A and

hypothetical protein, all of which arose independently through a different inactivating mutation, suggesting selection leading to loss of function in the bovine host (Guinane et al., 2010). Over 50 pseudogenes for genes involved in gene regulation, iron metabolism and virulence have also been observed in ovine *S. aureus* genomes (Le Maréchal et al., 2011).

Following the human-to-avian host jump resulting in the CC5 poultry-adapted *S. aureus* clone, loss of function of well-characterised genes involved in human disease pathogenesis was observed in the genome of avian ST5 strain ED98 (Lowder et al., 2009). In particular, all poultry ST5 *S. aureus* strains examined in this study had a premature stop codon in the Staphylococcal protein A gene (*spa*), which is known to be important in binding to human platelets and in evading phagocytosis mediated through non-specific binding of the Fc region of human immunoglobulin IgG (Foster, 2005; Hartleib et al., 2000). The avian equivalent of IgG known as IgY, has a structurally distinct Fc binding region that does not bind to SpA, which may explain why the loss of function of this gene occurred following transition to the avian host (Lowder et al., 2009; Warr, Magor and Higgins, 1995).

Taken together, these studies indicate that loss of gene function is a characteristic of *S. aureus* switching to a new ecological niche.

1.7.3 Adaptive allelic diversification

In response to selective pressures in different ecological niches, such as the host immune response or antibiotic use, *S. aureus* may acquire nonsynonymous mutations in genes that confer a fitness advantage, which may subsequently become fixed in the bacterial population (Guinane et al., 2010; McAdam et al., 2012; McCarthy and Lindsay, 2010). For example, the widespread use of fluoroquinolones is thought to have driven the evolution of resistance in *S. aureus* via several mechanisms, including the acquisition of mutations in the quinolone resistance determining regions (QRDRS) of genes *grlA/grlB* and *gyrA/gyrB* which encode the 2

fluoroquinolone target enzymes topoisomerase IV and DNA gyrase respectively (Takahashi et al., 1998). In addition, fluoroquinolone resistance can be mediated through the expression of efflux pumps (Kaatz, Seo and Ruble, 1993; Yoshida et al., 1990). In the CC30 clone, nonsynonymous mutations that are specific to the contemporary hospital-associated EMRSA-16 CC30 clones but not in the community-associated phage 80/81 and southwest pacific clones have been observed in *hla* encoding α-toxin, the global regulator Agr and the *crtM* gene involved in carotenoid biosynthesis, which may be the result of adaptation of the EMRSA-16 clone to the hospital environment (DeLeo et al., 2011; McAdam et al., 2012; McGavin, Arsic and Nickerson, 2012). These mutations likely lead to attenuated virulence, but may compensate for the metabolic cost of maintaining the multi-drug resistant phenotype of EMRSA-16 (McAdam et al., 2012; McGavin, Arsic and Nickerson, 2012).

In a study by McCarthy *et al*, core and core variable gene sequence variation was examined in *S. aureus* surface-expressed and secreted host immune evasion proteins (McCarthy and Lindsay, 2010). Variation was found to be lineage-specific, and largely concentrated in functional domains involved in host-pathogen interaction, suggesting that selective forces encountered are contributing to variation in these proteins (McCarthy and Lindsay, 2010). Although only a small number of animal strains were included in this study, the surface and immune evasion proteins were similar to those found in human strains, suggesting that these proteins are not essential for host specificity (McCarthy and Lindsay, 2010). However in a subset of surface and secreted proteins, either novel protein domain variants, or unique combinations of domain variants were identified in animal strains that were not present in human strains, indicating potential differing selective pressures in these different host environments (McCarthy and Lindsay, 2010).

For example, the *isdB* gene encoding the iron surface determinant protein B (IsdB) was one of several proteins found to have a unique combination of protein domain variants in animal strains that were not identified in human strains (McCarthy and Lindsay, 2010). The IsdB sequence of bovine strain RF122 also shows several

polymorphisms compared to human strains (Herron-Olson et al., 2007). IsdB is a surface-expressed haemoglobin receptor that is an essential enzyme for iron acquisition from haemoglobin, as the IsdB receptor captures haemoglobin on the *S. aureus* cell surface (Mazmanian et al., 2003; Torres et al., 2006). The IsdB receptor demonstrated enhanced specificity for binding human haemoglobin compared to haemoglobin from other mammals, although it was not determined whether animal-derived strains might have enhanced specificity for animal haemoglobin (Pishchany et al., 2010).

Selective pressures encountered in the ruminant host appear to have driven allelic diversification in additional genes that interact with the host (Zakour et al., 2008; Guinane et al., 2010). In a comparative genomic hybridisation study into the genome content of ruminant *S. aureus* compared to sequenced human strains, the majority of the regions of difference that were specific to ruminant hosts were found to be in core genome and not associated with MGE (Zakour et al., 2008). The majority of these were predicted to be in genes encoding extracytoplasmic proteins, suggesting a potential role in host-pathogen interaction or host specificity (Zakour et al., 2008). Considerable allelic variation, including higher than average dN/dS ratios were identified in virulence genes such as clumping factors A and B, and serine aspartate-repeat binding proteins, indicating that genes involved in host interaction in the ruminant host are under diversifying selection (Guinane et al., 2010; Herron-Olson et al., 2007).

1.8 *S. aureus* clonal complex 97 (CC97)

S. aureus CC97 is a major ruminant mastitis-associated lineage (Hata et al., 2010; Rabello et al., 2007; Smith et al., 2005). The first report of a CC97 mastitis strain in the literature is a single locus variant of ST97 (ST115), the Newbould 305 strain, which was isolated in 1958 from a case of bovine mastitis in Canada (Prasad and Newbould, 1968). Newbould 305 has been used as a reference strain for bovine mastitis infection studies (Atalla et al., 2009; Schukken et al., 1999; Shkreta et al., 2004).

S. aureus CC97 has been identified in ruminants worldwide (Figure 1.4). In a molecular typing study of over 250 bovine *S. aureus* isolates from the USA, UK and Chile, over 87 % of strains were identified as the CC97 clone (Smith et al., 2005). In Brazil, the genetic diversity of over 200 mastitis-associated *S. aureus* strains was assessed, with 91 % of strains represented by clonal lineages CC126 and CC97 (Rabello et al., 2007). *S. aureus* CC97 is also a major mastitis lineage in Asia, with a large-scale study of over 260 dairy farms over a 7 year period identifying that the majority of strains were represented by 2 lineages, CC97 and CC705 (Hata et al., 2010). In Europe, *S. aureus* CC97 is also reported as a major mastitis-associated lineage in several countries including Denmark (Hasman et al., 2010), the Netherlands (Ikawaty et al., 2009), Ireland (Smyth et al., 2009) and Germany (Monecke et al., 2007). The majority of isolates are from individual mastitis milk samples, or bulk tank milk (Hata et al., 2010; Ikawaty et al., 2009; Rabello et al., 2007), although CC97 has also been isolated from teat skin, milkers' hands and the environment, all of which may promote spread of contagious pathogens within the dairy parlour (Smith et al., 2005). CC97 appears to be associated with chronic subclinical mastitis, with histopathological changes observed in a mouse model of mastitis infection more consistent with chronic inflammation, in contrast to CC151, a lineage that appears to be associated with more severe infection, causing necrosis of the mammary tissues due to elevated expression of cytolytic toxins (Guinane et al., 2008).

Although the CC97 clone is associated with bovine mastitis, it has also been isolated from several other animal species, including other ruminants such as sheep, goats and buffalo (Aires-de-Sousa et al., 2007; Gharsa et al., 2012). In addition, pigs have been reported to be colonised with *S. aureus* CC97 in France (Lefevre et al., 2005), Italy (Battisti et al., 2010) and Germany (European Food Safety Authority (EFSA), 2009; Meemken et al., 2010). *S. aureus* CC97 has also been identified sporadically in rats (Himsworth et al., 2014; Van de Giessen et al., 2009), rabbits (Smyth et al., 2009), and an ostrich (Aires-de-Sousa et al., 2007). Most recently, a CC97 novel single locus variant, ST2712, was isolated from a wild-ranging roe deer in the Italian

alps, which are known to be in contact with domestic livestock in overlapping pastures (Luzzago et al., 2014).

The oldest known human-associated CC97 *S. aureus* is an MSSA from Denmark isolated in 1980 (Larsen, A.R. personal communication). The earliest report of CC97 *S. aureus* in humans is an MLST study of *S. aureus* isolates from Oxfordshire, UK, that were isolated in 1997 (Day et al., 2001; Feil et al., 2003). Since then, there have been increasing reports of CC97 isolated from humans worldwide, including North and South America (Diep, Perdreau-Remington and Sensabaugh, 2003; Mulvey et al., 2005; Schuenck et al., 2009), Africa (Ruimy et al., 2009, 2010), Europe including the UK (Ellington et al., 2008; Hallin et al., 2007; Krziwanek, Metz-Gercek and Mittermayer, 2011; Melles et al., 2004; Ruimy et al., 2009) and Asia (Chan et al., 2009; Peck et al., 2009; Sam et al., 2008). To date, there are reports in the literature citing human CC97 in over 35 different countries (Table 1.3). These range from nasal colonisation isolates (Ruimy et al., 2009), to skin infections such as cellulitis (Ellington et al., 2008), and more invasive life-threatening infections such bacteraemia (Vidal et al., 2009) and pleural empyema (Schuenck et al., 2009).

While there are no reports to date of ruminant-associated CC97 MRSA, many of the human-associated CC97 *S. aureus* reported in the literature are MRSA, with CA-MRSA SCC*mec* types IV (Vidal et al., 2009) and SCC*mec* type V (Ellington et al., 2008) most often reported. In a European-wide survey of CA-MRSA and MSSA involving 16 countries, 3 % of MRSA and 8 % of MSSA were found to be CC97 (Rolo et al., 2012). Human-associated CC97 has also been described in outbreak situations including an outbreak of ST97 CA-MRSA-V in a neonatal ward in Kuwait (Udo et al., 2011). These isolates were distinguished by ST and SCC*mec* type from other MRSA isolated from health-care workers or other patients in the same hospital, indicating an epidemiological link among the 15 affected neonates in the unit (Udo et al., 2011). An outbreak of CC97 CA-MRSA-V in the UK was reported previously (Ellington et al., 2008). Of note, the isolates contained ACME (See Section 1.5.2.3), indicating that this clone has acquired MGEs that may contribute to epidemic success.

Furthermore, there have also been sporadic reports of CC97 *S. aureus* that have been isolated from the environment in the USA, including a beach (Levin-Edens et al., 2011), the floor of a dental clinic (Roberts et al., 2011), and student housing (Roberts et al., 2011b).

Taken together, these findings prompt questions of what the evolutionary history of *S. aureus* CC97 is, and whether it is capable of human to human transmission. The genetic relatedness and basis for the wide host tropism of the animal and human CC97 *S. aureus* is currently unknown.

Figure 1.4. Map summarising the literature showing the countries and hosts in which *S. aureus* CC97 has been identified. The circles indicate countries in which *S. aureus* CC97 has been reported in the literature and is colour-coded according to the host species from which the strain was isolated. Blue circles are bovine isolates, green circles are pig isolates, and the red circles indicate human CC97 *S. aureus*.

Table 1.3. Countries in which human CC97 *S. aureus* have been identified

MLST CC (ST)*	Country	Reference
CC97	Algeria	(Ruimy et al., 2009)
CC97 (ST1796, ST97)	Argentina	(Lattar et al., 2012)
CC97 (ST97)	Australia	(Wehrhahn et al., 2010)
CC97 (ST97)	Austria	(Krziwanek et al., 2011)
CC97 (ST97)	Belgium	(Hallin et al., 2007)
CC97 (ST97)	Brazil	(Schuenck et al., 2009)
CC97 (ST97)	Canada	(Golding et al., 2011)
CC97 (ST669 - DLV/ST97)	Cape Verde	(Aires-de-Sousa, Conceição and de Lencastre, 2006)
CC97 (ST1419- SLV/ST97)	China	(Ho et al., 2010)
CC97 (ST97)	Denmark	(Bartels et al., 2007)

MLST CC (ST)*	Country	Reference
CC97	France	(Ruimy et al., 2009)
CC97 (ST97, ST1292 TLV/ST97)	French Guiana	(Ruimy et al., 2010)
CC97 (ST97)	Germany	(Strommenger et al., 2007)
CC97 (ST97)	Greece	(Chatzakis et al., 2011)
CC97 (ST97)	Iran	(Havaei et al., 2011)
CC97 (ST97)	Italy	Campanile, F. (http://www.mlst.net/)
CC97 (ST97)	Korea	(Peck et al., 2009)
CC97 (ST97)	Kuwait	(Udo et al., 2011)
CC97 (ST97)	Lebanon	(Tokajian et al., 2010)
CC97 (ST97, ST1179-SLV/ST97)	Malaysia	(Ghasemzadeh-Moghaddam et al., 2011)

MLST CC (ST)*	Country	Reference
CC97	Moldova	(Ruimy et al., 2009)
CC97 (ST97)	Netherlands	(Donker et al., 2009)
CC97	New Zealand	(Muttaiyah et al., 2010)
CC97 (ST97)	Norway	(Fossum Moen et al., 2009)
CC97 (ST97)	Portugal	(Conceição, Sousa and Lencastre, 2009)
CC97	Saudi Arabia	(Monecke et al., 2012)
CC97 (ST97)	Singapore	(Chan et al., 2009)
CC97 (ST97)	Spain	(Lozano et al., 2011)
CC97 (ST97)	Sweden	(Fang et al., 2008)
CC97	Switzerland	(Wattinger et al., 2012)
CC97 (ST97)	Tenerife	(Rivero-Pérez, Alcoba-Flórez and Méndez-Álvarez, 2012)
CC97 (ST97)	Tunisia	(Kechrid et al., 2011)

MLST CC (ST)*	Country	Reference
CC97 (ST97)	Turkey	Sudagidan, M (http://www.mlst.net)
CC97 (ST97)	United Kingdom	(Ellington et al., 2008)
CC97 (ST97)	United States of America	(Chung et al., 2004)

*SLV/ST97, single locus variant of ST97; DLV/ST97, double locus variant of ST97; TLV/ST97, triple locus variant of ST97.

1.9 Summary

Epidemiological surveillance and research into the basis by which zoonotic diseases may emerge or re-emerge is of utmost importance to public health worldwide. The interface between humans and animals provides ample opportunities for the continuing transmission and evolution of important pathogens such as *S. aureus* (Harrison et al., 2013; Van Loo et al., 2007). Recent studies employing high throughput sequencing have facilitated bacterial phylogenetic population studies, and highlighted the versatility and adaptability of *S. aureus* to different hosts and environments, with evidence that certain clones are continuing to adapt and expand into new hosts (Lowder et al., 2009; Price et al., 2012; Weinert et al., 2012). However, while many studies have been carried out on important human-associated MRSA clones, similar studies on animal-associated strains lag behind. This highlights the importance of continuing research into the molecular basis for the multiple host tropism of *S. aureus*, in order to identify emerging clones of public health importance, and to determine features of host adaptation that may be targeted by control measures and therapeutic interventions.

1.10 Project aims

The overall aim of this study is to investigate the molecular basis for the multiple host tropism of the major ruminant-associated clone, CC97 by:

- Determining the evolutionary history of CC97 *S. aureus*
- Identifying molecular correlates of host adaptation
- Identifying the processes contributing to adaptive genome diversification

2. Materials and Methods

2.1 Bacterial culture

Bacteria were streaked from frozen stocks onto tryptic soy agar (TSA) (Oxoid, UK) plates and incubated for 16 h at 37 °C or into tryptic soy broth (TSB) (Oxoid, UK) for overnight culture at 37 °C shaking at 200 rpm. Long term storage of strains was in 80 % (v/v) glycerol (Sigma-Aldrich, UK) at -80 °C .

2.2 Genomic DNA extraction and quantification

Overnight TSB culture was grown to stationary phase using the conditions described above (Section 2.1). Genomic DNA was isolated from a starting volume of 1 ml TSB culture using the PurElute^{TM} Bacterial Genomic Kit (Edge Biosystems, Maryland, USA) following the manufacturer's instructions, with a modification of the addition of 5 µl lysostaphin (5 mg/ml; AMBI products, NY, USA) to the cell lysis step. gDNA pellets were resuspended in 50 µl of Elution buffer (EB) (Qiagen, Manchester, UK). Alternatively, genomic DNA was isolated using the DNeasy Blood and Tissue DNA extraction kit (Qiagen, Manchester, UK) following manufacturer's instructions of the Bacterial Gram positive and yeast protocol using the QIAcube automated system (Qiagen, Manchester, UK).

The concentration of genomic DNA was quantified in ng/µl using a Nanodrop® ND-1000 Spectrophotometer (Nanodrop Technologies, DE, USA) to ensure a minimum total yield of at least 5 µg DNA for genome sequencing.

2.3 Polymerase chain reaction (PCR)

Unless otherwise stated, PCR reactions consisted of 40 ng genomic DNA (gDNA), 1 U GoTaq DNA Polymerase (Promega, Madison, USA), 100 nmol forward and

reverse primers (Life Technologies Ltd, Paisley, UK) (See Table 2.1), 5x GoTaq Flexi PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs (Promega, Madison, USA), and the remaining volume made up to a total of 25 µl with distilled water (dH_20). PCR cycling conditions were as follows: initiation step of 94 °C for 3 min, then 30 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 1 min, and a final extension step of 72 °C for 10 min. For a negative control, genomic DNA was replaced with the same volume of distilled water.

For reactions that required the use of a high fidelity polymerase enzyme, PCR reactions consisted of 25 µl volumes containing 40 ng gDNA, 1 U Pfu Ultra II Fusion DNA polymerase, 10 × Pfu Ultra II Reaction buffer (Agilent Technologies, CA, USA), 0.25 mM dNTPs (Promega, Madison, USA), 100 nmol forward and reverse primers (Life Technologies Ltd, Paisley, UK) (Table 2.1), with the remaining volume made up with dH_20. PCR cycle conditions were an initial step of 95 °C for 2 min, followed by 30 cycles of 95 °C for 20 s, 55 °C for 20 s and 72 °C for 15 s, with a final extension of 72 for 3 min.

Table 2.1. List of primer sequences used in this study

Primers	Sequence 5' to 3'	Size (bp)	Target
SasA-F	TCAACATCCTCAAAGAATACTACA	462	*sasA: sas* typing verification
SasA-R	ATGCGTTACTTAAGCCACCAATAC		
SasF-F	GGATAGCAAAGACAATAAAAGTTC	439	*sasF: sas* typing verification
SasF-R	TGATATGTGTAATGTTGCGTTGAG		
SasDfl-F	TGAGGTGGAGCATCTACAAGGTGTT	276 (*sasD* absent)	*sasD:* amplify different product sizes for *sasD* positive and negative isolates
SasDfl-R	AGGCTCGCTCAGATAATCACACGAT	1127 (*sasD* present)	
ArcC-F	TTGATTCACCAGCGCGTATTGTC	456	MLST: amplify *arcC* allele
ArcC-R	AGGTATCTGCTTCAATCAGCG		
AroE-F	ATCGGAAATCCTATTTCACATTC	456	MLST: amplify *aroE* allele
AroE-R	GGTGTTGTATTAATAACGATATC		

Primers	**Sequence 5' to 3'**	**Size (bp)**	**Target**
GlpF-F	CTAGGAACTGCAATCTTAATCC	465	MLST: amplify *glpF* allele
GlpF-R	TGGTAAAATCGCATGTCCAATT		
Gmk-F	ATCGTTTTATCGGGACCATC	429	MLST: amplify *gmk* allele
Gmk-R	TCATTAACTACAACGTAATCGT		
Pta-F	GTTAAAATCGTATTACCTGAAGG	474	MLST: amplify *pta* allele
Pta-R	GACCCTTTTGTTGAAAAGCTTAA		
Tpi-F	TCGTTCATTCTGAACGTCGTGA	402	MLST: amplify *tpi* allele
Tpi-R	TTTGCACCTTCTAACAATTGTA		
YqiL-F	CAGCATACAGGACACCTATTGGC	516	MLST: amplify *yqiL* allele
Yqil-R	CGTTGAGGAATCGATACTGGAAC		

Primers	**Sequence 5' to 3'**	**Size (bp)**	**Target**
Hlb-F	TTGATTTTGTTTATACCAAGCTTCA	541	Amplify intact β-haemolysin (*hlb*) gene
Hlb-R	GCCTGTAACTGTGTCAGAAGG		

2.4 Agarose gel electrophoresis

Electrophoresis of genomic DNA was carried out using a 0.8 % agarose gel in 1 × Tris-borate-EDTA (TBE) buffer (Sigma-Aldrich, Missouri, USA) containing 10 μg/ml SYBR safe DNA Gel stain (10 000 × concentrate in DMSO; Invitrogen, CA, USA), at 90 v for 90 min. Electrophoresis of PCR products was carried out in a 1.0 % agarose gel in 1 × TBE buffer containing 10 μg/ml SYBR safe DNA Gel stain, at 100 v for 45 min. DNA products were visualised with a MultiImage Light Cabinet (Alpha Innotech Corporation, USA).

2.5 PCR product purification

PCR products were purified using Exonuclease I-Antarctic Phosphatase (Exo-AP) digestion. A 1:1 ratio (25 μl each) of Exonuclease 1 and Antarctic Phosphatase (New England Biolabs, MA, USA) were combined and 44 μl of this mix was added to 22 μl 10x AP buffer (NEB) and 374 μl dH20. This ExoAP mix and PCR product were combined in a 1:1 ratio (8 μl each), and incubated at 37 °C for 15 min, 80 °C for 15 min followed by a hold step at 15 °C.

2.6 DNA Sanger sequencing

Sanger sequencing was carried out at The GenePool sequencing service (King's Buildings, University of Edinburgh, UK). The sequencing reaction included 2 μl purified DNA (3-10 ng template recommended for 200-500 bp PCR product, 5-20 ng for 500-1000 bp PCR product), 100 nM sequencing primer (See Table 2.1 for primer sequences), and 2 μl BigDye Terminator (Applied Biosystems, UK). Sequencing was carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit using manufacturer's instructions, using an ABI3730 capillary DNA analyser (Applied Biosystems, UK). Raw trace files were visualised and consensus sequence identified using Lasergene 8 software (DNAStar, Inc. Madison, WI, USA) (See Table 2.2).

2.7 Bioinformatics analysis

All software utilised during the course of this study and the general function of each program is summarised in Table 2.2. Where relevant, commandlines that were utilised as part of the study (preceded in each case by "$") and specific options that were implemented in order to analyse the dataset have been listed in the final column (Table 2.2). For the meaning of each option or flag, please refer to the relevant software manual (Table 2.2). The specific workflow pertaining to each chapter of the research project is described in Sections 3.3, 4.3 and 5.3.

Table 2.2 Software utilised during this study

Software	Version	Function(s)	URL/Source/Reference	Commandlines/Options specified
Lasergene	v8	Sequence editing	DNAStar, Inc. Madison, WI, USA. (Burland, 1999)	See manual for usage instructions.
FastQC	v0.10.0	Quality control for high throughput sequence data	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Babraham Institute, Cambridge, UK)	JAVA application. See manual for usage instructions.

FASTX toolkit	N/A	Quality pre-processing of short read FASTA/FASTQ files	http://hannonlab.cshl.edu/fastx_toolkit/ (Cold Spring Harbour Laboratory, New York, USA.)	$ fastx_quality_stats
				$ fastq_quality_boxplot_graph.sh
				$ fastx_nucleotide_distribution_graph.sh
				$ fastx_trimmer

Quake	v0.3.4	Detection and removal of sequencing errors in short reads	http://www.cbcb.umd.edu/software/quake/ (Kelley, Schatz and Salzberg, 2010)	$ count q-mers (-k15 flag)
				$ cov_model.py
				$ correct -f quake_text -k 15 -c 0.02 -m kmercounts_file -p 2

Software	Version	Function(s)	URL/Source/Reference	Commandlines/Options specified

BWA	v0.5.9	Short read alignment against reference genome	http://bio-bwa.sourceforge.net/ (Li and Durbin, 2009)	$ bwa index
				$ bwa aln
				$ bwa sampe

Samtools	v.0.1.16	Downstream processing of genome alignments	http://samtools.sourceforge.net/ (Li et al., 2009)	$ samtools view -b -S -q 30
				$ samtools sort
				$ samtools index
				$ samtools mpileup -E -Q 30 -q 30
				$ bcf tools view (-cg)

Velvet	v1.0.15	De novo assembly of short reads	www.ebi.ac.uk/~zerbino/velvet (Zerbino and Birney, 2008)	$ VelvetOptimiser.pl -s 19 -e 31 -f '-separate -shortPaired -fastq <input_files>' -t 4
Qualimap	v0.7.1	Quality analysis of alignment files (SAM/BAM)	http://www.qualimap.org (García-Alcalde et al., 2012)	See manual for usage instructions.
IGV	v2.0	Genome/Assembly visualisation	http://www.broadinstitute.org/igv/ (Robinson et al., 2011)	See manual for usage instructions.

Software	Version	Function(s)	URL/Source/Reference	Commandlines/Options specified
Artemis Comparison Tool (ACT)	N/A	Pairwise genome comparisons	http://www.sanger.ac.uk/resources/software/act/ (Carver et al., 2005)	See manual for usage instructions.
Galaxy	N/A	Multiple utilities for genome analysis	http://galaxyproject.org/ (Goecks, Nekrutenko and Taylor, 2010)	See manual for usage instructions.
BLAST tools (BLASTN/BLASTX/BLASTP)	N/A	Pairwise sequence comparisons (both DNA and amino acid sequences)	http://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul et al.,1990)	Online server or standalone version: $ blastn -query <query_file> -subject <subject_file> -out <output_file> -evalue 0.01 -outfmt 0
Mauve	v2.3.1	Multiple genome alignment and draft genome contig alignment	http://gel.ahabs.wisc.edu/mauve/ (Darling, Mau and Perna, 2010), (Rissman et al., 2009)	Progressive Mauve algorithm Mauve Contig Mover (MCM) tool
RDP	v4.13	Recombination detection in genome alignments	http://web.cbio.uct.ac.za/~darren/rdp.html (Martin et al., 2010)	Default settings. See manual for usage instructions.

Software	Version	Function(s)	URL/Source/Reference	Reference
BratNextGen	N/A	Bayesian analysis of recombination in whole genome samples	http://www.helsinki.fi/bsg/software/BRAT-NextGen/ (Marttinen et al., 2012)	• Draw proportion of shared ancestry tree (PSA tree) step (Hyper parameter alpha – set to 1, PSA tree cutoff set to 0.1) • Learning Recombinations step (30 iterations) • Estimating significance step (100 permutations, significance cutoff at 0.5)
Path-O-Gen	v1.3	Root-to-tip linear regression analysis of phylogenetic trees	http://tree.bio.ed.ac.uk/software/pathogen/ (Rambaut, A. Edinburgh University, UK)	See website for usage instructions.
BEAST	v1.7.1	Bayesian phylogenetic analysis	http://beast.bio.ed.ac.uk/ (Drummond and Rambaut, 2007)	• HKY model of sequence evolution with gamma correction for rate heterogeneity • Uncorrelated relaxed molecular clock model and 3-rate local clock model • Constant coalescent prior • Discrete trait evolution model used to predict ancestral state

Software	Version	Function(s)	URL/Source/Reference	Reference
RAxML	v7.2.6	Maximum likelihood phylogenetic analysis	http://sco.h-its.org/exelixis/web/software/raxml/index.html (Stamatakis, 2006)	$ raxmlHPC-PTHREADS-SSE3 -T 4 -f a -m GTRGAMMA -x 12345 -# 1000 -s <input_phylip_file> -n <project_name> -w <output_directory>

PAUP	v4.0b10	Calculation of consistency index	http://paup.csit.fsu.edu/ (Swofford, 2003)	$ paup> execute <input_nexus_file>;
				$ paup> cstatus;
				$ paup> tstatus;
				$ paup> describetrees 1/plot=phylogram brlens=yes;

MEGA	v4.0	Phylogenetic inference including the neighbour joining method	http://www.megasoftware.net/ (Tamura et al., 2007)	See manual for usage instructions.
Splitstree	v4.13.1	Phylogenetic network inference using molecular sequence data	http://www.splitstree.org (Huson and Bryant, 2006)/	• SNP alignment=input file • phi test implemented through this program

Software	Version	Function(s)	URL/Source/Reference	Reference
Cortex	v1.0.5.20	De novo assembly and analysis of genetic variation using coloured de Bruijn graphs	http://cortexassembler.sourceforge.net/index_cortex_var.html (Iqbal et al., 2012)	• Produce binary files: $ run_calls.pl --first_kmer 31 --fastaq_index fastq_index --auto_cleaning yes --bc yes --pd no --outdir dirname --outvcf casestudy2 --ploidy 1 --genome_size 3000000 --qthresh 5 --mem_height 21 --mem_width 100 --vcftools_dir <input_dir> --do_union yes --logfile log.txt --workflow joint --ref Absent --squeeze_mem • Analysis of genetic variation: $cortex_var_31_c48 --kmer_size 31 --mem_height 19 --mem_width 100 --colour_list pangenome_index --max_var_len 50000 --print_colour_coverages --print_novel_contigs 1/2/93/95/<output_name>

Software	Version	Function(s)	URL/Source/Reference	Reference
SnpEff	v3.0	Genetic variant effect prediction	http://snpeff.sourceforge.net/ (Cingolani et al., 2012)	$ java -jar ./snpEff.jar -c snpEff.config -a 20 -s snpEff_summary.html REF <input_vcf_file> > snpEff_output.txt
Prokka	v1.5.2	Prokaryotic genome annotation	http://vicbioinformatics.com/ (Seeman, 2014)	$prokka --outdir --addgenes --prefix -cpus 4 --mincontig 0 --gram pos <input_file>
RAST	v2.0	Online server rapid prokaryotic genome annotation	http://rast.nmpdr.org/ (Aziz et al., 2008)	Follow online server instructions.
EasyFig	v2.1	Creation of linear comparison of multiple genetic loci	http://easyfig.sourceforge.net/ (Sullivan, Petty and Beatson, 2011)	See manual for usage instructions.
BRIG	v0.95	Generation of circular prokaryotic genome comparison images using BLAST comparison	http://sourceforge.net/projects/brig/ (Alikhan et al., 2011)	See manual for usage instructions.

"$" precedes commandlines that were utilised in each program to analyse the CC97 datasets. For specific information relating to each of the options specified, please refer to the manual for that program, available at the URLs listed.

2.8 DNA Illumina genome sequencing

Paired end whole genome sequencing (WGS) of multiple indexed isolates was carried out using the sequencing service provided by ARK Genomics (Roslin Institute, University of Edinburgh, UK). Library preparation was carried out using either the Truseq DNA preparation kit or Nextera XT genomic kit (Illumina, San Diego, CA, USA), following the manufacturer's instructions. Sequencing to 36 bp or 250 bp read length was carried out with the Illumina Genome Analyzer IIx system or the Illumina MiSeq (Illumina, San Diego, CA, USA) following standard Illumina protocols (http://www.illumina.com).

Raw read FASTQ output files were assessed for quality and nucleotide distribution using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) (For commandlines used see Table 2.2) and the suite of quality analysis tools within FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). If necessary, reads were trimmed using the "fastx_trimmer" command within FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Mapped and *de novo* genome assemblies for each strain were constructed as described in Sections 3.3.2, and 4.3.2, using software listed in Table 2.2.

2.9 Comparative genomic analysis

BAM assembly files were visualised using IGV v2.0 (Robinson et al., 2011) to verify the quality of individual SNPs or genome regions if required, and to visualise coverage across different regions of the mapped genomes (Table 2.2). The quality of BAM files were analysed using Qualimap v0.7.1 (García-Alcalde et al., 2012) (Table 2.2). Comparison and visualisation of multiple genome sequences was made using Artemis Comparison Tool (http://www.sanger.ac.uk/resources/software/act/) (Carver et al., 2005) or Mauve v2.3.1 by aligning genomes using the progressive Mauve algorithm (Darling, Mau and Perna, 2010) (Table 2.2). Draft genome contigs were

aligned against a reference genome using the Mauve Contig Mover (MCM) tool available in Mauve (Rissman et al., 2009). Galaxy genome viewer (Goecks, Nekrutenko and Taylor, 2010) was utilised to perform basic file parsing to create suitable formats for software as listed in sections 3.3, 4.3 and 5.3 (See Table 2.2 for a full list of software).

3. Livestock Origin for a pandemic Community-associated MRSA clone

3.1 Introduction

Animals are a major source of new pathogens affecting humans (Woolhouse and Gaunt, 2007). However, the potential for pathogenic bacteria that originally were found in animals to switch hosts and become widely established in human populations is not clear. The results of population genetics studies have shown that most strains of *S. aureus* are host specific, indicating a low frequency of cross-species transmission (Fitzgerald, 2012). However, recent studies employing multilocus sequence typing (MLST) and whole-genome sequencing (WGS) have identified several *S. aureus* sequence types (ST) that are associated with multiple host species, implying either zoonotic transmission or a recent common ancestor. For example, the poultry-associated sequence type 5 (ST5) and livestock-associated, methicillin-resistant *S. aureus* (LA-MRSA) ST398 clones are descended from bacteria that recently made human-to-animal host jumps (Lowder et al., 2009; Price et al., 2012). Recently, Harrison *et al*. demonstrated that ST130 MRSA isolates containing the *mecC* gene have spread from cows to humans resulting in clinical infections (Harrison et al., 2013). However, the potential for animal-specialized strains of *S. aureus* to successfully cross the human species barrier and become epidemic in human populations is unclear.

A previous study by Weinert *et al.* (2012) reported that the CC59 *S. aureus* clone that is endemic in humans in Taiwan and has spread to other parts of the world may have originated in livestock about 500 years ago (Weinert et al., 2012), and this has recently been corroborated by Shepheard *et al.* (Shepheard et al., 2013). However, while MLST sequences are useful for examining the long-term evolution of *S. aureus*, it lacks resolution to identify much more recent host jump events leading to the emergence of new epidemic *S. aureus* clones (Weinert et al., 2012). *S. aureus* strains belonging to MLST CC97 are a leading cause of bovine mastitis in Europe, Asia, and North and South America (Smith et al., 2005; Smyth et al., 2009). Less commonly, CC97 has also been reported to cause infections of small ruminants, pigs, and humans (Battisti et al., 2010; Ellington et al., 2008; Smyth et al., 2009). Importantly, there have been increasing reports of human infections caused by CC97 isolates in Europe, North and South America, Africa, and Asia (Diep, Perdreau-

Remington and Sensabaugh 2003; Ruimy et al., 2009; Schuenck et al., 2009; Udo et al., 2011). However, the origin of human CC97 strains and their relatedness to livestock-associated CC97 strains are currently unknown. The aim of this study was to investigate the evolutionary history of the CC97 *S. aureus* lineage using genome sequencing technologies to determine the basis for the multiple host tropism of this clone.

3.2 Aims

- Determine the phylogenetic relationship for animal and human CC97 *S. aureus* to understand the genetic basis for multiple host tropism

- Utilise Bayesian phylogenetics to carry out temporal analysis to determine the mutation rate for CC97 and divergence times for any host-associated subclades identified

3.3 Materials and Methods

3.3.1 Bacterial isolates.

Over a period of several months, a total of 220 *S. aureus* CC97 isolates were obtained by requesting bacterial samples from the laboratories that had published details of strains available in the literature. These strains were isolated in 18 different countries on 4 continents, ranged in isolation date from 1956-2012, and the samples had been obtained from bovine, human, porcine and caprine hosts (Table 3.1). At the time of the study, it was possible to multiplex up to 12 isolates on a single lane of an Illumina slide. As such, the isolate genome collection was accumulated over time by sequencing isolates in batches of up to 12 strains maximum, as strains became available for use. Strains were selected from the larger collection to represent a range

of hosts, clinical, spatial and temporal variation (Table 3.1). Genomic DNA was isolated as described in Section 2.2. MLST was carried out to confirm the sequence type (ST) prior to whole-genome sequencing using methods described previously (Enright et al., 2000), and primer sequences as listed in Table 2.1 and the PCR conditions listed in Section 2.3. In total, although 48 isolates underwent whole genome sequencing, preliminary quality analysis (See Section 3.3.2) resulted in the omission of 5 isolates from the study due to reasons such as contamination, poor sequence quality, or errors with metadata, such as uncertainty regarding the date of isolation. This resulted in the current dataset of 43 isolates (Table 3.1).

Table 3.1: *S. aureus* strains used in this study

Study ID	Strain ID	Country (Region)	Host	Source	Date of isolation	ST (MLST)	SCC*mec* type	Reference
1_2008	12780	Italy	Porcine	Nasal swab	2008	97	MRSA-V	(Battisti et al., 2010)
2_2000	7231121_2	Denmark	Bovine	Milk-mastitis	2000	97	MSSA	DTU Food, Denmark
3_2008	7807	Italy	Porcine	Nasal swab	2008	97	MRSA-V	(Battisti et al., 2010)
4_2005	ALG33	Algeria	Human	Nasal swab	2005	97	MSSA	(Ruimy et al., 2009)
5_2003	CHILE5	Chile (Temuco)	Bovine	Milk-mastitis	2003	97	MSSA	(Smith et al., 2005)
6_2003	CHILE9	Chile	Bovine	Milk-mastitis	2003	97	MSSA	(Smith et al., 2005)
7_2003	CHILE22	Chile	Bovine	Milk-mastitis	2003	97	MSSA	(Smith et al., 2005)
8_2003	CO1899	UK	Bovine	Milk-mastitis	2003	97	MSSA	(Sung, Lloyd and Lindsay, 2008)
9_1987	CTH26	USA	Bovine	Milk-mastitis	1987	124	MSSA	(Smith et al., 2005)
10_1987	CTH54	USA	Bovine	Milk-mastitis	1987	97	MSSA	(Smith et al., 2005)

Study ID	Strain ID	Country (Region)	Host	Source	Date of isolation	ST (MLST)	SCC*mec* type	Reference
12_1987	CTH163	USA	Bovine	Milk-mastitis	1987	97	MSSA	(Smith et al., 2005)
13_2007_83	07-02708	Turkey	Human	Unknown	05/10/2007	97	MSSA	(Strommenger et al., 2007)
14_2006_58	06-01420	Germany (Lubeck)	Human	Unknown	05/07/2006	97	MRSA-IV	(Strommenger et al., 2007)
15_2006_75	Fli5576-9	Germany (Thuringia)	Bovine	Unknown	27/09/2006	Novel SLV (*glpF*)	MSSA	(Monecke et al., 2007)
16_2005	FRA134	France	Human	Nasal swab	2005	Novel SLV (*aroE*)	MSSA	(Ruimy et al., 2009)
17_2008	H068N	French Guiana	Human	Nasal swab	2008	97	MSSA	(Ruimy et al., 2010)
18_1997	H118	UK	Human	Blood	1997	28	MSSA	(Enright et al., 2000)
19_2005	HO54060126	UK (West Midlands)	Human	Pressure sore	2005	97	MRSA-V	(Ellington et al., 2008)

Study ID	Strain ID	Country (Region)	Host	Source	Date of isolation	ST (MLST)	SCC*mec* type	Reference
20_2005	HO54120523	UK (West Midlands)	Human	Pressure sore	2005	97	MRSA-V	(Ellington et al., 2008)
21_2006	HO64020492	UK (West Midlands)	Human	Groin swab	2006	97	MRSA-V	(Ellington et al., 2008)
22_2007	HO70420647	UK (West Midlands)	Human	Wound swab	2007	97	MRSA-V	(Ellington et al., 2008)
23_2007	HO71540648	UK (West Midlands)	Human	MRSA screening swab	2007	97	MRSA-V	(Ellington et al., 2008)
24_2009	HO91460210	UK (West Midlands)	Human	Cellulitis of leg	2009	97	MRSA-V	(Ellington et al., 2008)
25_2009	HO91740182	UK (East England)	Human	Skin abscess	2009	97	MRSA-V	(Ellington et al., 2008)
26_1999_33	3177	UK	Human	Blood	13/04/1999	97	MSSA	(Feil et al., 2003)
27_1956	JVI215	Denmark	Bovine	Mastitis	1956	97	MSSA	DTU Food, Denmark
28_2005_91	563	Brazil	Human	Pleural fluid-empyema	11/2005	97	MRSA_IV	(Schuenck et al., 2009)

Study ID	Strain ID	Country (Region)	Host	Source	Date of isolation	ST (MLST)	SCC*mec* type	Reference
29_1964	LMA1166	France	Bovine	Mastitis	1964	97	MSSA	(Zakour et al., 2008)
30_2005	MOLD123	Moldova-Chisinau	Human	Nasal swab	2005	97	MSSA	(Ruimy et al., 2009)
31_2007	KK24	Turkey (Izmir)	Human	Blood	2007	97	MRSA-IV	http://www.mlst.net
33_1958	NCIMB702892	Canada	Bovine	Milk	1958	115	MSSA	(Smith et al., 2005)
34_1996	NZ189	New Zealand	Human	Wound swab	1996	97	MSSA	(Zinn, Westh and Rosdahl, 2004) (SARISA/Westh, H).
35_2007	SA5	Spain (Asturias)	Bovine	Milk-mastitis	2007	97	MSSA	http://www.mlst.net
36_1993_83	RF111	Ireland	Bovine	Milk-mastitis	1993	97	MSSA	(Fitzgerald et al., 1997)
37_1993_91	RF115	Ireland	Bovine	Milk-mastitis	10/11/1993	97	MSSA	(Fitzgerald et al., 1997)
38_1993_91	RF116	Ireland	Bovine	Milk-mastitis	10/11/1993	97	MSSA	(Fitzgerald et al., 1997)
39_2007	B1020	UK (Brighton)	Human	Eye swab	2007	97	MRSA-IV	(Miller et al., 2010)

Study ID	Strain ID	Country (Region)	Host	Source	Date of isolation	ST (MLST)	SCC*mec* type	Reference
40_2007	SA8	Spain	Bovine	Milk-mastitis	2007	352	MSSA	http://www.mlst.net
41_1996	UC030	USA	Human	Surgical ward	1996	97	MSSA	(Zinn, Westh and Rosdahl, 2004) (SARISA/Westh, H).
42_2003	VETBG42	Portugal	Caprine	Milk	2003	97	MSSA	de Lencastre, H.
43_1980	10668	Denmark	Human		1980	97	MSSA	Larsen, A. Statens Serum Institut
44_2007	55435	Denmark	Human		2007	97	MRSA	Larsen, A. Statens Serum Institut
45_2011	76325	Denmark	Human		2011	97	MRSA	Larsen, A. Statens Serum Institut
46_2012	83075	Denmark	Human	Blood	2012	97	MRSA	Larsen, A. Statens Serum Institut

3.3.2 Genome sequencing, mapping assembly and SNP calling.

Paired-end Illumina sequencing of bacterial strains was carried out on an Illumina genome analyzer IIx or Miseq, following standard Illumina protocols. Nucleotide distribution and quality scores of raw reads were assessed using FastQC v0.10.0 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and filtered for quality using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) (Table 2.2). High quality reads were aligned against the reference genomes of *S. aureus* strains MW2 (accession number NC_003923) or Mu50 (accession number NC_002758.2), using the Burrows-Wheeler short aligner v0.5.9 (BWA), implementing the BWA-SW (Smith Waterman) algorithm (Li and Durbin, 2009) (For details of commandlines and options used see Table 2.2). Consensus sequences were called and variants (point mutations and small insertions and deletions) detected for sites with a minimum read depth of 3, and minimum PHRED scores and average mapping values above 30, with extended BAQ computation employed in Samtools v.0.1.16 (Li et al., 2009) (Table 2.2). The core genome was defined as all nucleotide sites shared by all isolates and putative recombinant regions were removed after breakpoint detection with the suite of programs included in the Recombination Detection Program v4.13 (RDP) (Martin et al., 2010). For comparisons of gene content, *de novo* assemblies of short reads was carried out using Velvet v1.0.15 (Zerbino and Birney, 2008) and the VelvetOptimiser.pl script within VelvetOptimiser v-2.1.7 (http://bioinformatics.net.au/software.velvetoptimiser.shtml) (Table 2.2).

3.3.3 Phylogenetic analysis

Core genome and in-frame protein coding sequences were extracted from the genome consensus sequences to construct alignments using custom scripts. To extract in-frame CDS, the python script utilised the genome coordinates of each CDS listed in the Genbank file of the reference genome, to extract the nucleotides at the

same coordinates from the mapped assemblies of each strain. These sequences were then filtered to remove CDS with up to a user-defined percentage of uncalled bases to remove CDS that were unmapped or only partially mapped. In this study, any CDS with more than 10 % uncalled bases was defined as "unmapped" and filtered out of the sequence files for all strains in the dataset. The script to extract core genome examined each column of the whole mapped genome alignment for all strains in a step-wise manner. For those sites at which a base was called in all strains, the base and genome coordinate were written to new output files. Sites with one or more uncalled base in the column were passed over and not written to the output file. This resulted in a core genome alignment containing only those sites shared by all strains included in the dataset, along with a list containing the genome coordinates included in the alignment.

The maximum likelihood phylogeny was reconstructed using RAxML-7.2.6 (Stamatakis, 2006) implementing a GTR model with gamma correction for rate heterogeneity, and 1000 bootstrap replicates (Table 2.2, Figure 3.1). To assess the phylogeny for temporal signal and estimate the validity of assuming a molecular clock to the data, linear regression analysis of root-to-tip distances was conducted on the maximum likelihood phylogeny using Path-O-Gen v1.3 (http://tree.bio.ed.ac.uk/software/pathogen/). Bayesian phylogenetic analysis (performed by P.R. McAdam) was conducted using BEAST v1.7.1 (Drummond and Rambaut, 2007) implementing the Hasegawa-Kishino-Yano model of sequence evolution with gamma correction for rate heterogeneity (Table 2.2). Rates were calculated by the dated tip method using a 3-rate local clock model, with the livestock rate constrained to 1.53×10^{-6} nucleotide substitutions per site per year as determined using an uncorrelated relaxed molecular clock model with a constant coalescent prior (Drummond et al., 2006) (Table 2.2). For the date of the MRCA with the human outgroup ST28, an in-frame coding sequence (CDS) alignment of CC97 and ST28 sequences was employed which allowed the estimation of mutation rate over all sites and third codon positions only. In each case, the mutation rates were comparable and resulted in similar estimates of the host jump date (Table 3.2). Ancestral host states were predicted using a model of discrete trait evolution (Lemey

et al., 2009), adapting the method by substituting geographic location with host state as described previously (Weinert et al., 2012). The consistency index for parsimony informative sites was determined using PAUP v4.0b10 (Swofford, 2003) (Table 2.2).

3.3.4 Antimicrobial sensitivity testing.

All isolates in the study were tested using the Vitek 2 system (Analysis done by G. McAllister) (AST-P620 card) (BioMérieux UK Limited, Basingstoke, UK) with a panel of antimicrobial agents including cefoxitin, benzylpenicillin, oxacillin, gentamicin, ciprofloxacin, inducible clindamycin resistance, erythromycin, clindamycin, linezolid, daptomycin, teicoplanin, vancomycin, tetracycline, tigecycline, nitrofurantoin, fusidic acid, mupirocin, chloramphenicol, rifampicin and trimethoprim. Results for antibiotic susceptibility were interpreted using standards defined by the Clinical and Laboratory Standards Institute (CLSI) (Clinical and Laboratory Standards Institute, 2007).

3.3.5 Statistical analysis of human *S. aureus* data from Denmark.

As part of the study by Spoor *et al*., collaborators Anders Larsen and Robert Skov (Statens Serum Institut, Denmark) were approached to conduct statistical analysis on their extensive dataset of human MRSA and bacteraemia infections in Denmark, to determine whether there is an increasing prevalence of human CC97 cases (Spoor et al., 2013). The reason for this is that uniquely, in Denmark, submission of all MRSA to the Danish National MRSA Reference Laboratory, Statens Serum Institut, has been mandatory since November 2006, and all MRSA and bacteraemia *S. aureus* isolates submitted since 2007 have been genotyped by *spA* typing. As this includes all strains, and not just pathogenic clones of interest that have been sent to the laboratory, an increase in prevalence in a specific *spa* type is a true representation of an increase in prevalence against the background of all other clones, and therefore

not under any form of sampling bias. As such, the collaborators were able to analyse *spa* types associated with CC97 from 2007 to 2011, on which the Cochran-Armitage trend test was employed to determine if there was an increase in prevalence of CC97 *spa* types over the 5 year period (Analysis performed by A.R. Larsen) (GraphPad Prism 5.0). In total, 11838 *S. aureus* strains were included in the analysis, consisting of 7109 *S. aureus* bacteraemia cases and 4729 MRSA. Of these, 124 were CC97 bacteraemia strains and 58 were CC97 MRSA. A selection of these human CC97 strains were included for whole genome sequencing in the current study, specifically strains 43_1980, 44_2007, 45_2011 and 46_2012 (Table 3.1).

3.3.6 Accession Numbers

The Illumina sequences generated and used in this study are deposited and available in the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under the study accession number PRJEB1411, located at http://www.ebi.ac.uk/ena/data/view/PRJEB1411. The *S. aureus* isolates are under sample accession numbers ERS212248 to ERS212287 and ERS249844 to ERS249847.

3.4 Results

3.4.1 CC97 *S. aureus* is an emerging cause of human infections

In order to investigate the possibility that infections due to CC97 strains are increasing in prevalence in human populations, as part of the study by Spoor *et al.*, the number of CC97-associated strains isolated from MRSA and bacteraemia infections in Denmark between 2007 and 2011 were determined by collaborators in Denmark on the extensive dataset of mandatory *S. aureus* bacteraemia and MRSA

submissions (analysis conducted by Anders R. Larsen, see Section 3.3.5) (Spoor et al., 2013).

In Denmark, cases of MRSA caused by CC97-related *spA* types increased from a total of 2 in 2007 to 22 in 2011, equivalent to an 11-fold increase in 5 years. This represents a significant increase in prevalence from 0.3 % to 1.7 % of total annual MRSA in Denmark since 2007 ($P < 0.01$).

3.4.2 Human ST97 *S. aureus* clones originated from bovine-to-human host jumps

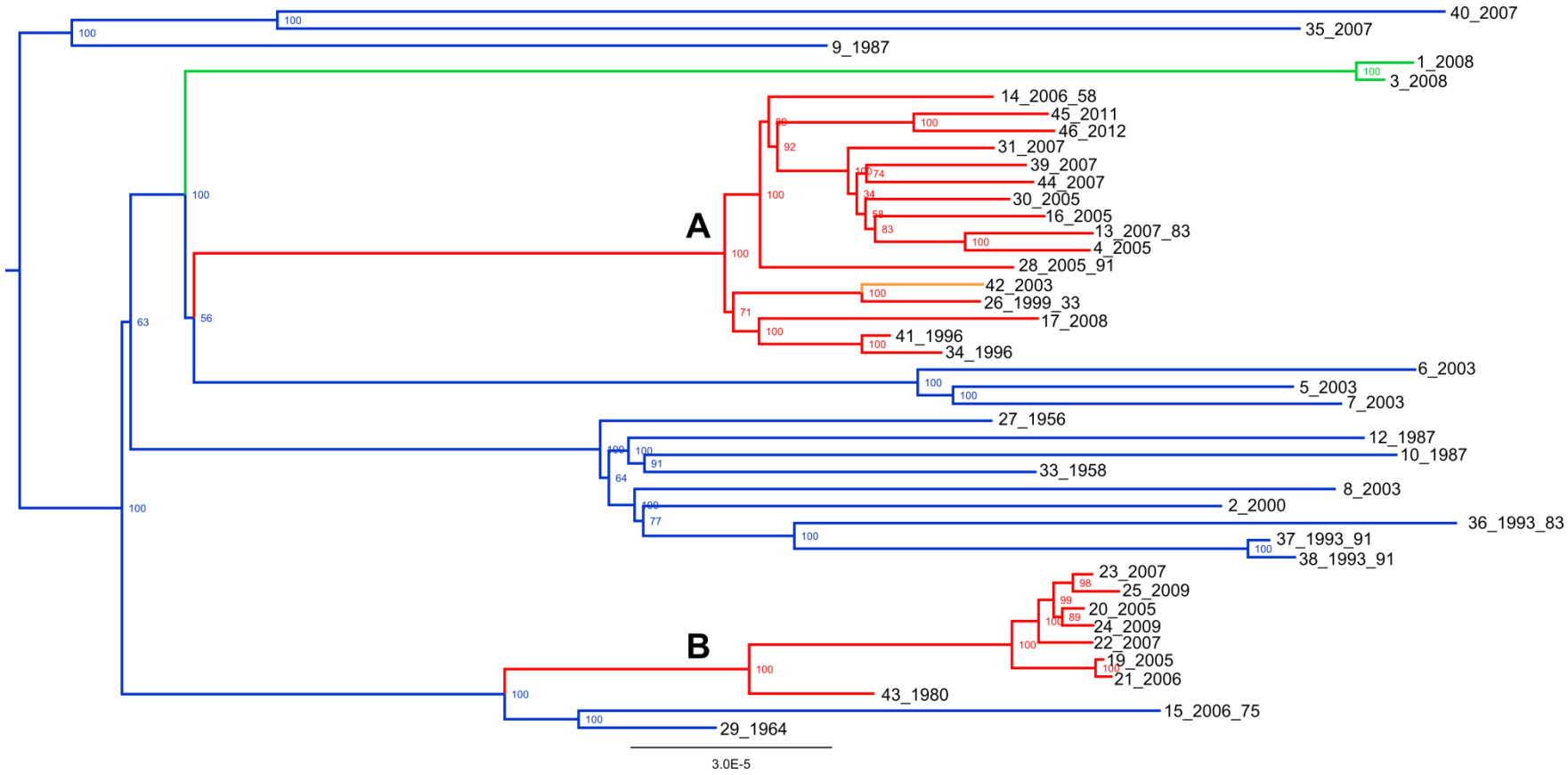
In order to investigate the evolutionary history of the CC97 clone, 220 CC97 *S. aureus* isolates of human, bovine, porcine, and caprine origin were obtained, isolated in 18 countries on four continents between 1956 and 2012. For whole-genome sequencing, 43 CC97 *S. aureus* isolates (including 16 MRSA) were selected, which broadly represented the breadth of host, geographic, and temporal variation identified among strains reported in the literature, in addition to a single isolate of the closely related ST28 as a human-specific outgroup. After putative recombinant regions had been stripped out of the alignment, the core genome of the 43 CC97 isolates consisted of 2 079 972 bp which included 5425 high-quality single nucleotide polymorphisms (SNP).

A maximum likelihood tree was constructed from the core genome alignment, resulting in a phylogeny with strong bootstrap support at most nodes (Figure 3.1). Overall, the high-resolution phylogenetic tree resolves *S. aureus* CC97 into distinct host-associated clades (Figure 3.1). Of note, there is considerable genetic diversity among CC97 isolates of livestock origin, which is indicated by numerous deep branches in the phylogenetic tree, compared to human isolates which are restricted to two distinct clades of closely related isolates (Figure 3.1). The majority of livestock-associated CC97 isolates lie basal to the human clades, indicative of a livestock

origin for the human strains (Figure 3.1). Human CC97 clade A consists of isolates of both MSSA and MRSA from 12 different countries on four different continents, indicating its global dissemination (Figure 3.1). In contrast, human CC97 clade B is represented by a single MSSA isolate from a Danish patient in 1980 and several MRSA bacteria isolated within the last 10 years in the Midlands region of the United Kingdom, indicating a more limited geographic distribution among the strains sampled (Ellington et al., 2008).

To determine the temporal signal of the CC97 phylogeny, and assess the suitability of inferring a phylogeny under a molecular clock model, an exploratory root-to-tip linear regression analysis of the dated-tip maximum likelihood tree was performed, plotting root-to-tip distance of each taxa on the phylogenetic tree against sampling date (Figure 3.2). The R^2 value was 0.3522, the correlation coefficient was 0.5978 and the residual plot is as shown in Figure 3.2 (B). The R^2 value is similar to the R^2 value of 0.27 previously published for similar analysis conducted on the ST225 *S. aureus* (Nubel et al., 2010). The slope of the graph and estimated evolutionary rate was 1.2821×10^{-6}, consistent with previously estimated mutation rates for other *S. aureus* clones (Harris et al., 2010; McAdam et al., 2012; Nübel et al., 2010) (Figure 3.2).

Figure 3.1 (See next page). Maximum likelihood phylogenetic tree of CC97 *S. aureus* lineage based on core genome alignment rooted with an outgroup strain of ST28 (18_1997). Branches are colour-coded according to host species of isolation (blue, bovine; green, porcine; orange, caprine; red, human, labelled clade A and B as shown). The scale bar represents the number of nucleotide substitutions per site. Bootstrap values (based on 1000 replicates) are shown for each node.


40_2007
35_2007
9_1987
1_2008
3_2008
14_2006_58
45_2011
46_2012
31_2007
39_2007
44_2007
30_2005
16_2005
13_2007_83
4_2005
28_2005_91
42_2003
26_1999_33
17_2008
41_1996
34_1996
A
6_2003
5_2003
7_2003
27_1956
12_1987
10_1987
33_1958
8_2003
2_2000
36_1993_83
37_1993_91
38_1993_91
23_2007
25_2009
20_2005
24_2009
22_2007
19_2005
21_2006
B
43_1980
15_2006_75
29_1964
3.0E-5

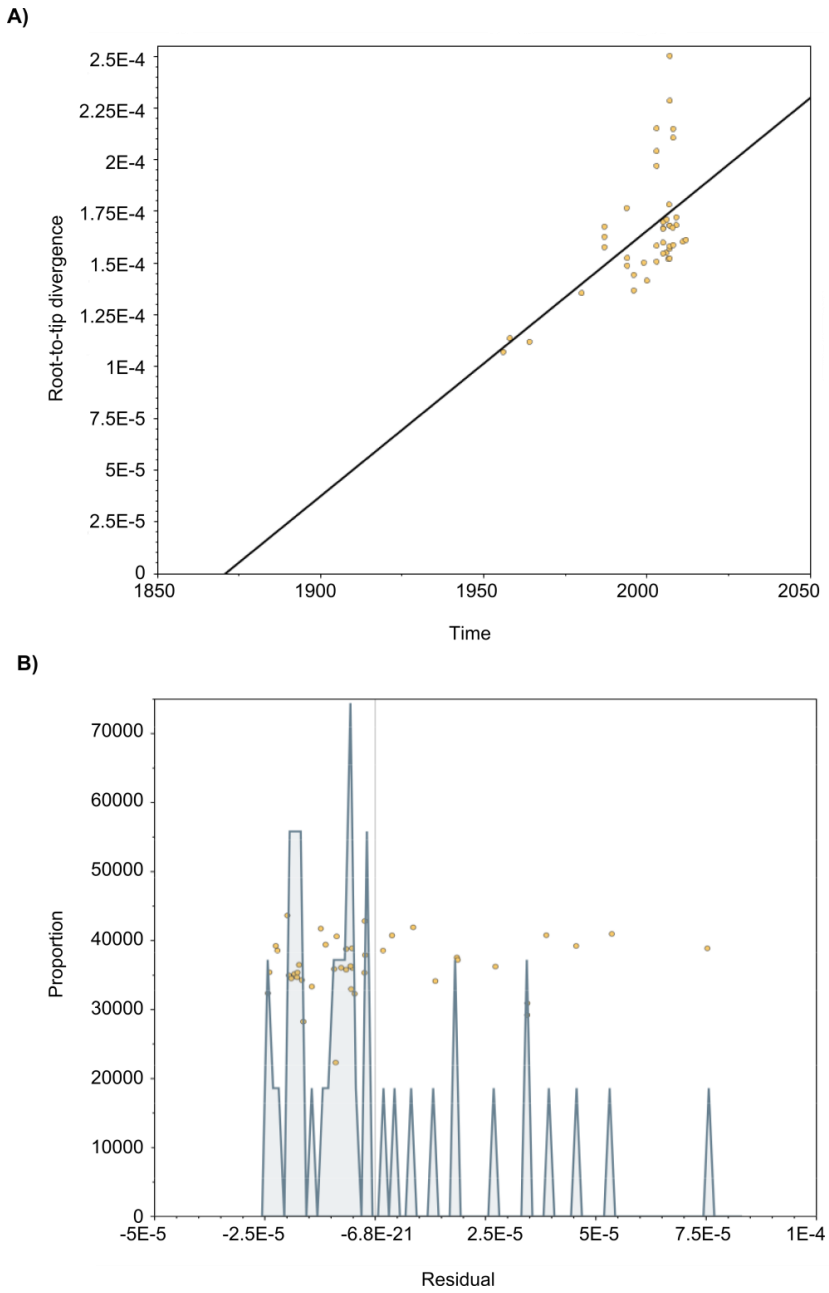


Figure 3.2: Linear regression analysis of root-to-tip distances of maximum likelihood CC97 *S. aureus* phylogeny using Path-O-Gen v1.4. A) Linear regression analysis plotting time of isolation against the root-to-tip distances on the phylogenetic tree. There is a positive correlation of divergence against sampling date. B) Residual plots to determine whether the observed error is stochastic and to examine for bias within the dataset. Although the data is biased to some degree for isolates of a more recent date, the temporal signal overall was deemed sufficient to proceed with Bayesian phylogenetic analysis.

3.4.3 CC97 *S. aureus* livestock-to-human host jumps are estimated to have occurred at least 40 years ago

To estimate the evolutionary rate of the CC97 lineage, and infer the estimated timing of each livestock-to-human host jump, additional phylogenetic reconstruction of the CC97 lineage was carried out using the BEAST program (named BEAST for Bayesian evolutionary analysis by sampling trees) (Drummond et al., 2012) (Figure 3.3) (Analysis in BEAST conducted by Lucy A. Weinert and Paul R. McAdam). The consistency index for parsimonious sites predicted using PAUP*4.0b10 (Swofford, 2003) indicated a very low predicted frequency of homoplasic events (consistency index [CI] = 0.95), implying an unconflicted phylogenetic signal.

In order to determine the time frame of the livestock-to-human host jump events, the mutation rates for the CC97 lineage were determined, allowing for variation in rates associated with different clades. Initially, an uncorrelated lognormal relaxed molecular clock model was used to determine a mutation rate for the livestock strains only of 1.53×10^{-6} nucleotide substitutions per site per year (95 % highest posterior densities [HPDs] of 1.35×10^{-6} to 1.72×10^{-6}). This rate was then fixed for livestock strains, and a local rate clock model was applied to each of the human clades, resulting in estimates of 9.58×10^{-7} (7.80×10^{-7} to 1.15×10^{-6}) for human clade A and 1.29×10^{-6} (1.05×10^{-6} to 1.54×10^{-6}) for human clade B. The Bayesian analysis resulted in estimates of host jump events which occurred between 1894 and 1977 for human CC97 clade A and between 1938 and 1966 for human CC97 clade B (Figure 3.3, Table 3.2).

In addition, the date of the most-recent human ancestor (MRCA) of CC97 and the human outgroup ST28 was estimated to be AD 784 (BC 325 to AD 1460), which lies within the credibility intervals of the previous estimate of the most recent minimum date for the host jump event using MLST data (Table 3.2) (Weinert et al., 2012).

3.4.4 Methicillin resistance was acquired by human CC97 clones subsequent to the host jumps from livestock

Antimicrobial susceptibility testing of CC97 isolates (VITEK analysis conducted by Gina McAllister) revealed that 7 of 17 bovine isolates were sensitive to all antimicrobial agents tested, with a further 6 resistant to only a single agent (Figure 3.4). In contrast, 20 of 23 human isolates were resistant to at least one antimicrobial agent, with some strains resistant to β-lactam antimicrobials, lincosamides, erythromycin, and trimethoprim (Figure 3.4).

Importantly, methicillin resistance is a characteristic of the human CC97 clades, with two distinct staphylococcal cassette chromosome *mec* element (SCC*mec*) types, types IV and V, which are associated with human CC97 clades A and B (Figure 3.3). Specifically, 14 of the 23 human CC97 *S. aureus* contained the SCC*mec* element, and 2 of 2 porcine CC97 isolates contained SCC*mec* (Figure 3.3). In human clade A, 5 strains had SCC*mec* type IV, and 2 strains isolated in 2011 and 2012 in Denmark, had SCC*mec* type V (Figure 4.1 SCC*mec* type column). In contrast, all 7 of the clade B UK CA-MRSA strains and both pig isolates contained SCC*mec* type V (Figure 3.3). Of note, none of the isolates examined contained the novel *mecC* allele previously found among several bovine MRSA clones responsible for episodes of human zoonotic infection (Harrison et al., 2013).

In contrast, all of the bovine *S. aureus* isolates examined in the study are methicillin sensitive, suggesting that resistance was acquired after the host jump from cows to humans. Consistent with this, the earliest human isolate identified in clade B (isolated in Denmark in 1980) was sensitive to all antibiotics tested, whereas all other isolates of clade B (isolated since 2005) were resistant to multiple antimicrobial classes (Figure 3.4).

Recently, it was demonstrated that methicillin and tetracycline resistance was likely acquired by LA-MRSA ST398 strains by antibiotic selective pressures encountered within the pig farming industry (Price et al., 2012). Of note, the two ST97 isolates from pigs in the current study were resistant to both tetracycline and methicillin, in addition to ciprofloxacin (Figure 3.4). Overall, the antimicrobial susceptibility profiles of the human and pig CC97 isolates demonstrated resistance to a much greater number of antimicrobials than the bovine CC97 *S. aureus* (Figure 3.4).

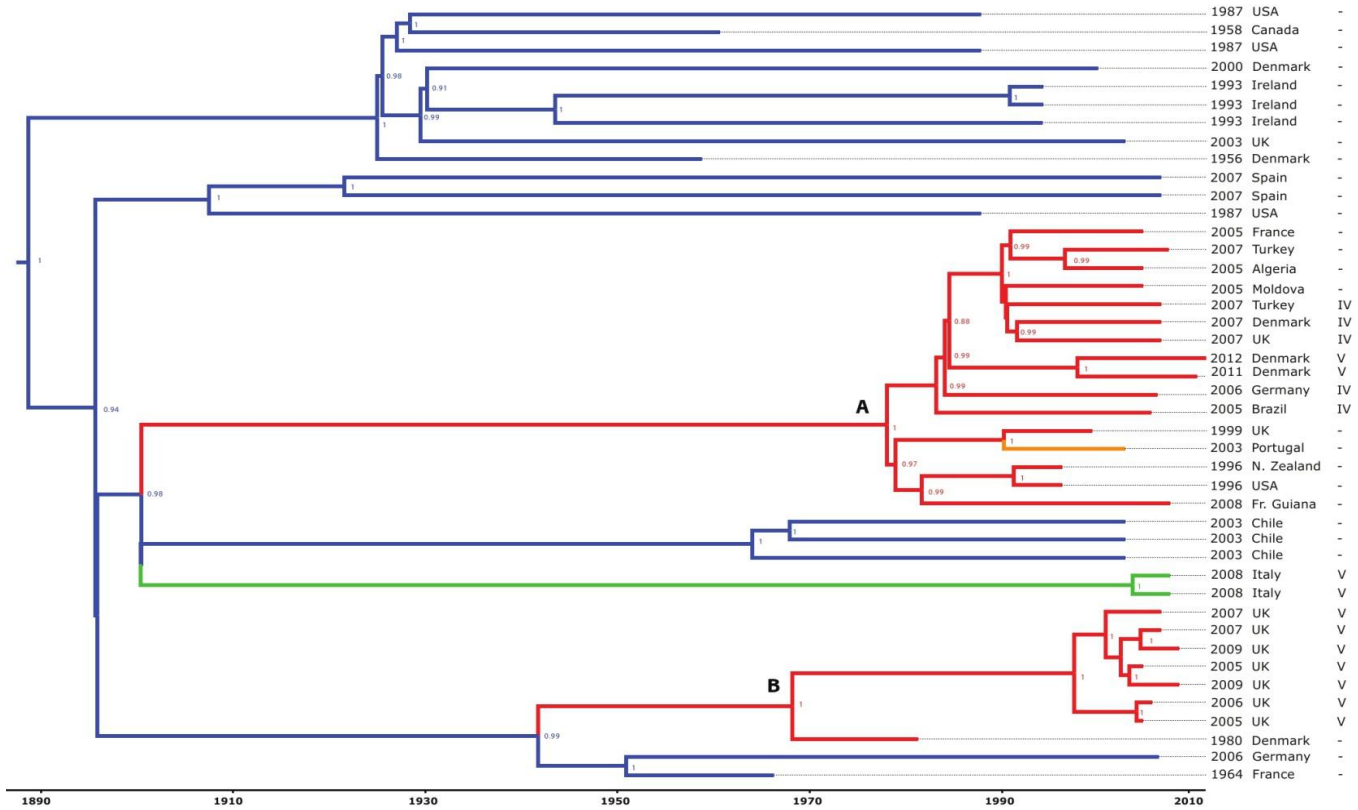


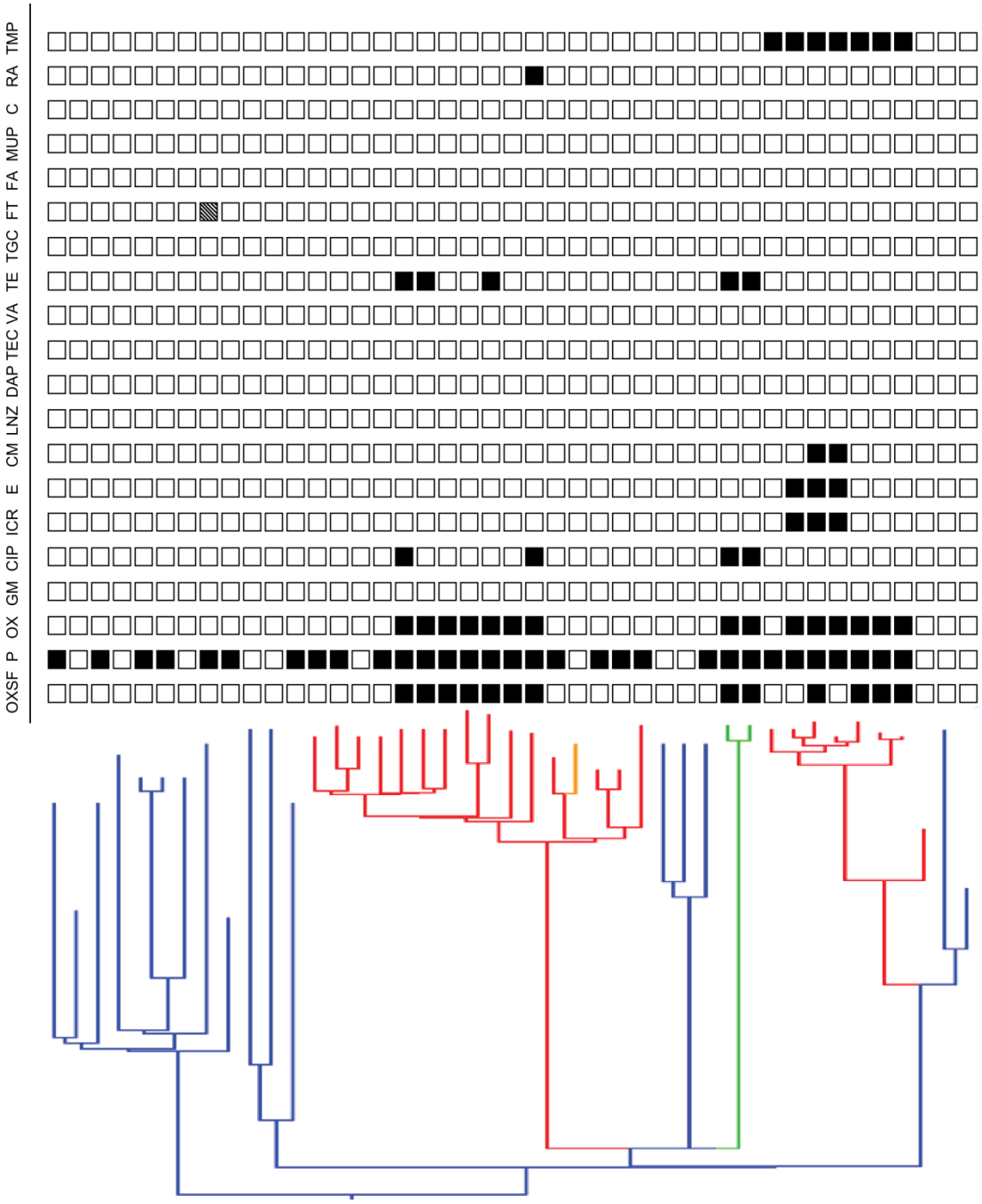
Figure 3.3. Identification of human *S. aureus* clones which resulted from livestock-to-human host jumps. Bayesian phylogenetic reconstruction of the CC97 lineage based on core genome alignment. Branches are colour-coded according to host species (blue, bovine; green, porcine; orange, caprine; red, human), and date and country of origin of each isolate indicated. MRSA SCC*mec* type (IV or V) is indicated on the right of the diagram. Branch lengths are scaled according to years. (BEAST analysis by Paul R Mcadam).

Table 3.2. Estimated dates of host jump events in CC97 *S. aureus* lineage

Host jump event	**Date of ancestral node Mean (95 % HPD)**	**Date of MRCA Mean (95 % HPD)**	**Ancestral host state (probability)**
Livestock-to-human (clade A)	1894 (1880-1907)	1977 (1972-1980)	Bovine (0.91)
Livestock-to-human (clade B)	1938 (1930-1944)	1966 (1961-1970)	Bovine (0.90)
Human-to-livestock (CC97)	784 (325 BC-1460)	-	-

(BEAST analysis conducted by Paul R McAdam).

Figure 3.4. Antimicrobial sensitivity profiles of CC97 *S. aureus* (See next page). The antibiogram for each strain is shown as a horizontal panel against the corresponding taxa as depicted in the Bayesian phylogeny. Sensitivity to each antimicrobial based on MIC data is indicated as follows: S, susceptible (white boxes); R, resistant (black boxes); I, intermediate resistance (hatched box); Antimicrobial abbreviations are as follows: OXSF, cefoxitin; P benzylpenicillin; OX, oxacillin; GM, gentamicin; CIP, ciprofloxacin; ICR, inducible clindamycin resistance; E, erythromycin; CM, clindamycin; LNZ, linezolid; DAP, daptomycin; TEC, teicoplanin; VA, vancomycin; TE, tetracycline; TGC, tigecycline; FT, furantoin; FA, fusidic acid; MUP, mupirocin; C, chloramphenicol; RA, rifampicin; TMP, trimethoprim. (Antimicrobial susceptibility testing using VITEK 2 card P260 conducted by Gina McAllister).


TMP
RA
C
MUP
FA
FT
TGC
TE
VA
TEC
DAP
LNZ
CM
E
ICR
CIP
GM
OX
P
OXSF

3.5 Discussion

Although CC97 *S. aureus* is recognised as a dominant clone associated with bovine mastitis (Hata et al., 2010; Smith et al., 2005), the genetic basis for the ability of this clone to colonise other hosts had not yet been established. In the current study, epidemiological data from Denmark over several years as analysed by collaborators as part of the study by Spoor *et al.*, revealed a significantly increasing trend in human cases of *S. aureus* CC97 MRSA (Spoor et al., 2013). Although equivalent detailed epidemiological data is not available for other countries, since the first human CC97 isolate was reported in 1996 (Westh, Zinn, and Rosdahl, 2004), there have been increasing reports of human *S. aureus* CC97 isolated from nasal carriage and clinical isolates, with CC97 now being identified in over 35 countries (Table 1.3). A recent study of community-associated *S aureus* isolated from humans in 16 countries in Europe reported that 3 % of CA-MRSA and 8 % of CA-MSSA were ST97 isolates (Rolo et al., 2012). CC97 has also been identified as the second most prevalent *S. aureus* genotype in a study on 94 osteomyelitis patients in Argentina (Lattar et al., 2012). Taken together, these data suggest that CC97 is an emerging cause of human infections.

The CC97 dataset size of the current study at 43 strains, is smaller than some more recent bacterial genome sequencing studies, which often sequence over 100 *S. aureus* genomes (Castillo-Ramirez et al., 2012; Holden et al., 2013; Everitt et al., 2014), while a study by Croucher *et al*., has sequenced 240 strains of *Streptococcus pneumoniae* (Croucher et al., 2011). This is reflective of the higher cost of sequencing of bacterial strains at the time the current CC97 study began (2010). Over time, the cost of sequencing multiple bacterial genomes has decreased, partly due to the fact that multiplexing of isolates on a single Illumina lane has increased from 12 to 96, thereby reducing costs associated with the number of sequencing runs per study and the amount of reagents required. However, the number of strains included in genome sequencing studies is highly variable. For example, studies into *S. aureus* ST239 population structure and host association of EMRSA-15 in humans and

companion animals employed datasets of 63 and 46 strains respectively (Harris et al., 2010; Harrison et al., 2014). Previous population genomic studies of *Clostridium difficile* and *Chlamydia psittaci* have used 30 and 20 strains respectively (He et al., 2010; Read et al., 2013). In the current study, the statistical support at the nodes of the Bayesian and maximum likelihood phylogenetic trees indicates that the sample size is adequate and the conclusions drawn in this study are supported. While every effort was made to reduce sampling bias, the nature of the study, in which isolates were sourced from published strains in the literature means that selection bias cannot be completely eliminated, as one must rely on others for sampling of strains from different countries, in addition to their availability. For example, it was not possible to obtain isolates from Asia, even though there are both human and bovine CC97 strains published in the literature (Peck et al., 2009; Hata et al., 2010; Ho et al., 2010). In addition, the quality and variability of metadata varies depending on the source of the isolate. However, where possible, isolates were selected to represent as much variation as possible in terms of host, date of isolation and geographical location.

The high resolution afforded by high throughput sequencing enabled the CC97 clone to be resolved phylogenetically, identifying distinct host-adapted sublineages. The current study identified the evolutionary origins of the human *S. aureus* CC97 as originating from livestock. The phylogenetic analysis indicates that this is the result of at least 2 independent livestock-to-human host jumps (Figure 3.1, Figure 3.3). Domestic animals have previously been identified as the source for the emergence of human pathogens such as the measles virus which is predicated to have evolved from rinderpest virus in cattle (Furuse et al., 2010), and swine-origin influenza A virus (S-OIV) H1N1, which arose from reassortment of swine viruses circulating in domestic pigs (Smith et al., 2009). However the current study is the first using whole genome sequences to identify recent host jump events by *S. aureus* from livestock into humans.

Previously, genomic studies have identified host jump events for *S. aureus* from humans into animals. For example, the major poultry-associated ST5 clone was shown to have arisen from a single human-to-poultry host jump over 40 years ago, originating from a human ST5 strain which may have been circulating in or near Poland (Lowder et al., 2009). A recent phylogenetic study by Price *et al.* (2012) identified human ST398 MSSA as the origin for CC398 LA-MRSA (Price et al., 2012). Evolutionary genomic analysis of the small ruminant clone CC133 estimates that this clone arose as a result of a human-to-livestock host jump, with evidence in the ruminant genome of adaptive diversification, including gene decay and the acquisition of novel host-specific mobile genetic elements (Guinane et al., 2010). Genomic features within *S. aureus* ST151 strain RF122 also indicates this bovine strain originated from a human ancestor, such as the decay of genes known to be important in human pathogenesis, including protein A and clumping factor A (Herron-Olson et al., 2007). Sakwinska *et al.* (2011), also identified a subtype of the widespread human genotype ST8 which is endemic to cows in Western Switzerland, indicating a possible human origin (Sakwinska et al., 2011).

Previous studies using MLST sequences have identified human clones CC59 and CC25 as having had livestock-associated ancestors, with CC59, which is endemic in Taiwan, estimated to have emerged from livestock approximately 500 years ago (Shepheard et al., 2013; Weinert et al., 2012). In comparison, the CC97 clone has jumped from livestock to humans much more recently, with estimates that this may have occurred between 32 and 40 years ago (Table 3.2). The livestock-to-human host jumps identified in the current study may reflect a wider phenomenon occurring in other *S. aureus* clones. Until now, the limited resolution of MLST has precluded rigorous examination of the occurrence of more-recent livestock-to-human host jump events leading to the emergence of new epidemic *S. aureus* clones (Weinert et al., 2012). In recent years, MRSA containing a novel *mecC* allele has been identified in bovines and humans (García-Álvarez et al., 2011; Shore, et al., 2011), primarily in the ruminant-associated clones CC130 and ST425, indicating a possible bovine reservoir of infection. The *mecC* MRSA has also been found to colonise multiple

host species, including humans, companion animals and wildlife (García-Álvarez et al., 2011; Paterson et al., 2012; Walther et al., 2012). Although prevalence of human infections varies among some European countries, in Denmark, there was evidence of increasing numbers of human infections of *mecC* positive MRSA between 2003 and 2011 according to Denmark epidemiological data (Petersen et al., 2013). While there is evidence of isolated zoonotic transmission events occurring (Harrison et al., 2013), as yet, host jump events followed by epidemic spread in humans have not yet been identified in the *mecC* positive MRSA, although some isolates have been found in humans without livestock contact, including one case of human-to-human household transmission (García-Garrote et al., 2014). Therefore, further phylogenetic studies will be required to determine if there are any recent host-jump events to account for the wide host tropism of this clone and the increasing infections identified in humans, and for the early identification of other emerging animal-associated clones in human populations.

Phylogenetic analysis of the evolutionary origins of the CC97 lineage as a whole suggest that CC97 emerged from a human-associated ancestor approximately 1200 years ago. This estimate lies within the credibility intervals of the previous estimate of the most recent minimum date for the human-to-livestock CC97 host jump event using MLST data (Weinert et al., 2012). The estimated time of emergence for CC97 is well after the domestication of cattle approximately 10 000 years ago, although estimates indicate that other ruminant lineages such as CC151/CC130 and CC133/CC425 have been associated with ruminants for longer, emerging approximately 5000 years ago and 3000 years ago, respectively (Weinert et al., 2012). However, taken together, these data indicate that CC97 is a well-established ruminant lineage that has been co-evolving with ruminants for hundreds of years, and has subsequently undergone recent host jumps back into humans.

In the current study mutation rates for the dataset were estimated using clock models implemented in BEAST. The estimated mutation rate for livestock of 1.53×10^{-6}

nucleotide substitutions per site per year (95 % highest posterior densities [HPDs] of 1.35×10^{-6} to 1.72×10^{-6}), and of 1.29×10^{-6} (1.05×10^{-6} to 1.54×10^{-6}) for human clade B are consistent with rates previously identified for other *S. aureus* clones. Previous mutation rates estimated for other clones of *S. aureus* range from 1.42×10^{-6} for CC30 (McAdam et al., 2012) to 5.125×10^{-6} for ST5 (Lowder et al., 2009). The human clade A strains have a slower estimated mutation rate of 9.58×10^{-7} (7.80×10^{-7} to 1.15×10^{-6}), suggesting that mutation rates within a single lineage can vary, which may reflect variations in selective pressures encountered in different ecological niches. For example, if the strains of human clade A were not well adapted to the human ecological niche and had decreased fitness relative to other CC97 strains, this may have had downstream effects on generation time. Alternatively if at some point there was a bottleneck or decrease in effective population size, this might result in a bacterial subpopulation with extremely low genetic diversity.

Human CC97 clade A consists of isolates of both MSSA and MRSA from 12 different countries on four different continents, indicating that it has undergone global dissemination since the host jump into humans occurred. The strains within human clade A are differentiated by 924 SNPs, which is much lower than the 3685 SNPs which differentiate the bovine strains, consistent with the more recent emergence of the human clade. In a study by Nubel *et al*, very low genetic diversity was discovered in the European clade of ST225 *S. aureus* compared to the genetically diverse established US ST225 clone, with spatio-temporal dynamic analysis indicating that the European ST225 clone emerged from a single transmission from the US, followed by rapid expansion across Europe (Nübel et al., 2010). The mutation discovery method examined approximately 4 % of the genome at 11884 bp, with the European ST225 found to have 41 bi-allelic polymorphisms, which the authors estimate as a pairwise genome comparison of approximately 28 SNPs, suggesting a lower level of genetic diversity than the human clade A strains (Nübel et al., 2010).

In contrast, human CC97 clade B is represented by a single MSSA isolate from a Danish patient in 1980 and several MRSA bacteria isolated within the last 10 years in the Midlands region of the United Kingdom, representing a more limited geographical distribution. This clade is differentiated by 225 SNPs, and the 7 UK CA-MRSA in this clade are more representative of an outbreak, given their limited range in date of isolation from 2005-2009, their restricted geographic location and low genetic variability (Ellington et al., 2008). An outbreak of CC97 *S. aureus* has been observed previously in a neonatal unit in Kuwait, further highlighting the ability of this strain to transmit successfully between humans (Udo et al., 2011). However, given the more recent estimated host jump event leading to the emergence of human clade B compared to human clade A, the lower genetic variability may represent an early stage of clonal expansion in this subclade, so further sampling of strains over time will be required in order to determine whether this subclone continues to expand or whether it has been confined to a UK outbreak.

Methicillin resistance is a characteristic of the human CC97 clades, with two distinct staphylococcal cassette chromosome *mec* element (SCC*mec*) types, types IV and V, which are associated with human CC97 clades A and B (Figure 3.3). The independent acquisition of multiple types of SCC*mec* within a single clone has also been observed in other human *S. aureus* clones such as ST5 (Nübel et al., 2008), CC8 (Robinson and Enright, 2003) and CC22 (Boakes et al., 2011). In contrast to the human strains, all of the bovine *S. aureus* isolates examined in the study are methicillin sensitive, suggesting that resistance in the human strains was acquired after the host jump from cows to humans. Consistent with this, the earliest human isolate identified in clade B (isolated in Denmark in 1980) was sensitive to all antibiotics tested, whereas all other isolates of clade B (isolated since 2005) were resistant to multiple antimicrobial classes. Recently, it was demonstrated that methicillin and tetracycline resistance was likely acquired by LA-MRSA ST398 strains by antibiotic selective pressures encountered within the pig farming industry (Price et al., 2012). Of note, the two ST97 isolates from pigs in the current study were resistant to both tetracycline and methicillin, in addition to ciprofloxacin.

Of note, 2 of the human clade A strains that carried SCC*mec* type V were the 2 most recent isolates from this clade (Figure 3.3). This may represent a shift in carriage from SCC*mec* type IV to SCC*mec* type V, however further study requiring additional CC97 strains would be required to verify this. The SCC*mec* type V element in the CC97 pig strains and human clade B strains have the *czrC* gene encoding cadmium and zinc resistance protein C (Cavaco et al., 2010), which has been identified previously on SCC*mec* type Vc elements (Li et al., 2011). The methicillin resistance of these strains may be co-selected for by the presence of zinc in pig feed as previously postulated for CC398 strains (Aarestrup et al., 2010; Cavaco et al., 2011; Price et al., 2012).

In contrast, bovine CC97 *S. aureus* strains did not contain SCC*mec* with either *mecA*, nor the variant resistance gene *mecC* (García-Álvarez et al., 2011; Harrison et al., 2013). These findings suggest that SCC*mec* was acquired subsequent to the host jump into humans, rather than cattle being a host for MRSA that was then transferred to humans. In general the CC97 bovine strains were resistant to fewer antimicrobials in comparison to the pig and human strains, despite the widespread use of antimicrobials such as intramammary preparations in the dairy industry (VMD, 2012).

Overall, the antimicrobial susceptibility profiles of the human and pig CC97 isolates demonstrated resistance to a much greater number of antimicrobials than the bovine CC97 *S. aureus* (Figure 3.4). These data imply that the dairy industry does not strongly promote the emergence of antibiotic-resistant *S. aureus,* in spite of the widespread use of antibiotics for treating bovine mastitis. The difference in levels of antimicrobial resistance may reflect differences in farming practices between the dairy and pig industries, which would affect transmission dynamics and alter selective pressures that the bacteria encounter. For example, in intensive pig rearing facilities, growing and finishing pigs may be kept at high stocking densities, with

antimicrobials often administered therapeutically or prophylactically as medicated feeding stuff (MFS) to groups of housed pigs to prevent the spread of pathogens (http://www.vmd.defra.gov.uk/ProductInformationDatabase/). Indeed, the size of breeding and production units has been implicated as a risk factor for MRSA colonisation of pigs, with the risk for nasal colonisation increasing with the number of breeding pigs (EFSA, 2010). The antimicrobial resistance genes contained in the CC97 *S. aureus* pig isolates, which are both from Italy, are likely to reflect antimicrobial selective pressures in Italian pig farming. For example, consistent with the tetracycline resistance genes and phenotype observed in both CC97 pig strains, tetracycline resistance in *S. aureus* isolated from pigs appears to be widespread in Italy and other European countries such as the Netherlands (De Neeling et al., 2007; Van Duijkeren et al., 2008). The widespread use of long-acting third and fourth generation injectable cephalosporins in pig farming has also been implicated in selecting for fluroquinolone-resistant bacteria capable of zoonotic transmission, such as extended spectrum beta-lactamase(ESBL)-producing Enterobacteriaceae (Battisti et al., 2010; European Medicines Agency, 2009; Jørgensen et al., 2007). In order to gain a wider understanding of the resistance profiles of CC97 *S. aureus* in pigs, antibiograms from an additional panel of CC97 pig isolates from other countries should be screened.

In contrast, in the dairy industry, although there is geographical variation in husbandry practices based on climate, the herd will spend a proportion of the year grazing pasture, with transmission of *S. aureus* occurring between cows within the milking parlour, the opportunity for which arises twice daily or more, notwithstanding biosecurity measures that may be in place. Dairy cows are selectively bred to produce high milk yields and in the weeks post-calving are under a considerable amount of metabolic stress which may affect their ability to fight infection. Another hypothesis for persistence of *S. aureus* mastitis infection status within herds is that in chronic subclinical mastitis infections, the bacteria occupy an intracellular niche within bovine mammary epithelial cells that also protects them from antimicrobial selective pressures (Garzoni and Kelley, 2009). Taken together,

these findings demonstrate that livestock are a potential source of MSSA in humans, that can subsequently acquire methicillin resistance and disseminate globally. The recent observation of the CC130 lineage containing the variant *mecC* allele as a cause of zoonotic transmission events (Harrison et al., 2013), taken together with the data from the current study highlights the importance of further surveillance measures to identify emerging pathogenic clones of livestock origin that may adapt to become successful human clones of public health importance. Overall, the data from the current study demonstrates that livestock are one potential reservoir of pathogenic bacteria with the capacity to cross the species barrier, undergo host-adaptive evolution, and become established in human populations, and as such are a potential route for emergence that might explain the recent global increase in some clones of CA-MRSA.

The importance of hygiene in prevention of hospital transmission of nosocomial pathogens such as MRSA is widely appreciated and the recent reduction in hospital MRSA infections is likely due in part to improved hygiene measures for controlling transmission (Wyllie et al., 2011). The data from the current study suggest that interactions between livestock and humans can promote the transmission of *S. aureus* clones with the potential to become epidemic in humans. Improved biosecurity and hygiene control measures which prevent the spread of bacterial flora between livestock and human hosts may limit opportunities for successful livestock to human transmission. Furthermore, regular surveillance of the microbiota in livestock and humans may facilitate the early identification of emergent clones with the capacity to transmit and cause disease among human populations. Further work will be to expand this analysis to other *S. aureus* lineages that colonise multiple hosts to determine the frequency of host jump events occurring in the wider *S. aureus* population.

4. Molecular correlates of host-association in *S. aureus* CC97

4.1. Introduction

Previous studies have revealed genome diversification of *S. aureus* which may contribute to host adaptation (Guinane et al., 2010; Herron-Olson et al., 2007; Lowder et al., 2009). For example, genetic mutations resulting in loss of gene function were observed in the genome of ruminant strain RF122, in genes encoding products known to be important in human disease pathogenesis, such as Staphylococcal protein A (SpA) and clumping factor A (ClfA), consistent with adaptation to the ruminant host from a human-associated ancestor (Herron-Olson et al., 2007). In comparison to human CC5 *S. aureus*, the genome of avian ST5 strain ED98 was also found to contain a number of pseudogenes that would encode proteins involved in human disease pathogenesis, including SpA, and Accessory secretory protein 1 (Asp1), which is required for expression of SraP, a cell wall-associated protein that is involved in human platelet adhesion and pathogenic in infective endocarditis (Lowder et al., 2009).

Adaptive genome diversification may also involve the acquisition of MGE encoding products which confer a potential fitness advantage in transitioning to a new ecological niche. For example, examination of the complement of MGE in the poultry CC5 clone indicated that following the human-to-poultry host jump, poultry CC5 *S. aureus* strains acquired avian-specific MGE such as plasmids, prophages and SaPIs, implying an important role for these MGE in adaptation to the avian host environment (Lowder et al., 2009). In addition, the small ruminant strain ED133 (CC133) was shown to have acquired a complement of ruminant-specific MGE, including a SaPI encoding a von Willebrand factor binding protein (vWbp) that was demonstrated to be responsible for coagulating ruminant plasma, thereby contributing to ruminant-specific *S. aureus* pathogenesis (Guinane et al., 2010). The β-converting phage, containing the IEC genes that are known to encode products important in evasion of human innate immune defences is found in over 90 % of human *S. aureus* strains (Wamel et al., 2006), while less than 20 % of ruminant *S. aureus* are reported to have this phage (Monecke et al., 2007; Sung, Lloyd and

Lindsay, 2008), indicating that the presence of this phage may confer a fitness advantage in the human host.

Selective pressures that are encountered in certain environments, such as antimicrobial use or the host immune response, may result in the selection of beneficial mutations associated with adaptive diversification by *S. aureus*. Allelic diversification in genes that encode proteins known to interact with the host has been observed in ruminant *S. aureus* (Zakour et al., 2008; Guinane et al., 2010). For example, nonsynonymous mutations were observed in surface-expressed genes encoding ClfA and ClfB in the genomes of ruminant strains ED133 and RF122 that were not present in gene sequences from human strains, and elevated dN/dS ratios were identified in several genes known to be involved in host interactions, suggestive of diversifying selection (Guinane et al., 2010; Herron-Olson et al., 2007).

Given the recent livestock-to-human host jumps that have been identified in the CC97 *S. aureus* lineage (Chapter 3), the aim of this part of the study was to examine the CC97 strains for genetic evidence of host adaption to the human host.

4.2 Study Aims

- Examine the distribution of MGE between the CC97 strains isolated from cows, pigs and humans that may reflect adaptation to different host species

- Investigate SNPs and indels associated with host in CC97 *S. aureus* that may contribute to host adaptation

4.3 Materials and Methods

4.3.1 Bacterial isolates and genome sequences

S. aureus CC97 strains isolated from cows, humans and pigs that were analysed in this study were the same strains as listed in Table 3.1. Bacterial growth conditions are as described in Section 3.3.1. The genomic DNA and genome sequences used for analysis are the same as those described in Section 3.3.2. In this study, comparative genomic content analysis of the CC97 *S. aureus* strains was analysed using a combination of genome assembly sequences, including reads mapped to reference genome MW2 (Accession number NC_003923) and the *de novo* Velvet assemblies constructed using Velvet v1.0.15 (Zerbino and Birney, 2008) with the VelvetOptimiser.pl script within VelvetOptimiser v-2.1.7 (http://bioinformatics.net.au/software.velvetoptimiser.shtml) (Table 2.2).

4.3.2 Analysis of variation in MGEs among CC97 *S. aureus* strains from different hosts

Comparative genomic analysis of variation in MGEs was conducted using Cortex1.0.5.20 (Iqbal et al., 2012), which utilises coloured de Bruijn graphs to construct *de novo* assemblies and detects variant sequence among defined groups within bacterial population datasets (Iqbal et al., 2012) (Table 2.2). Here, paired end Illumina FASTQ reads were used as input. Genetic variation was assessed by defining groups of CC97 *S. aureus* strains by host (cow, human and pig), then comparing assembly contigs using Cortex to define sequences that were present in 1 host group but not in the other host groups (Iqbal et al., 2012) (Table 2.2). Protein coding sequences (CDS) were annotated using Prokka -1.5.2 (Prokka: Prokaryotic Genome Annotation System (http://vicbioinformatics.com/) or RAST (Aziz et al., 2008), in addition to BLASTX query against the nr Genbank database (Altschul et

al.,1990) (Table 2.2). For characterisation of MGE, the appropriate Velvet *de novo* contigs were examined to identify MGE CDS and insertion sites.

4.3.3 Identification of SNPs and indels associated with host in CC97 *S. aureus*

Paired end Illumina FASTQ reads were mapped to the ST1 reference genome MW2 (NC_003923) as described in Section 3.3.2 (Table 2.2). Variants were called using SAMtools v.0.1.16 (Li et al., 2009) (Table 2.2). Variant VCF files were used as input for the software SnpEff v 3.0 (Cingolani et al., 2012) which predicts the effect of each variant, identifying nonsynonymous SNPs, synonymous SNPs, intergenic variants, frameshift mutations and premature stop codons (Cingolani et al., 2012) (Table 2.2). SnpEff results for all strains were collated together, from which non-core variants and non-informative variants (variants present in all strains against the reference) were filtered out. Variants at coordinates identified as specific to each clade of the CC97 *S. aureus* phylogenetic tree (See Section 3.4.2) were identified using custom shell scripts. This was done using the linux "awk" command to run through the list of effects and identify how many times each coordinate was listed, and sort the results by number. For example, to identify the non-informative coordinates listed in all 43 strains, the following command was used:

$ cat <snpEff_input_file> | cut -f1,3 | sort -n +1 -2 | uniq -c | awk '{print $3}' | sort -n | uniq -c | sort -n -r +0 -1 | awk '$1=="43"{print $0}' > <output_file >.

These output files were further filtered by strain identification using "awk" to identify coordinates specific to each phylogenetic clade. For genes in which clade-associated variants were identified, all other variants detected in that gene were examined manually, and the variants were designated host-associated if there were no other effects identified in that gene from strains of other host species.

4.4 Results

4.4.1 The accessory genome of human CC97 *S. aureus* has diversified since the host switch from cows

In order to determine molecular determinants of *S. aureus* CC97 associated with host species, variation in MGE distribution was examined in strains isolated from bovine, human and pig host species.

4.4.1.1 Bacteriophages

The family of β-converting phages (φSa3) associated with human *S. aureus* contain an immune evasion cluster (IEC) of genes encoding secreted proteins such as staphylokinase (SAK), staphylococcal complement inhibitor (SCIN), and chemotaxis inhibitory protein of *S. aureus* (CHIPS) which contribute to immune evasion in a human host-specific manner (Wamel et al., 2006). In the current study, 19 of 23 human isolates contained a φSa3 with an IEC (Figure 4.1). In contrast, none of 19 bovine or pig isolates contained φSa3 (Figure 4.1). Of the 16 human CC97 clade A isolates, 14 had an IEC containing genes *sak* and *scn* (IEC type E), and of the 8 human CC97 clade B isolates, 4 had an IEC containing *sak*, *chp*, and *scn* (IEC type B) and one isolate had IEC type E (Figure 4.1) (Wamel et al., 2006). Of note, the single goat strain that co-segregates with human clade A contains a φSa3 phage which has 99 % nucleotide identity to the φSa3 phage in human clade A strains (Figure 4.1).

Comparative analysis of representative sequences for the 2 IEC phage types (type E and type B) revealed that approximately 55 % of the phage content is homologous, demonstrating 98 % nucleotide identity across 25.3 kbp of the phage, with approximately 20 kbp unique to each phage type (Figure 4.2A). The majority of

regions of difference are within the central region of the phage, with 23 CDS specific to the IEC type E sequence, 26 CDS specific to the IEC type B sequence, and 37 CDS shared between the 2 phage sequences (Figure 4.2A).

In addition, strain 45_2011, a Danish human CC97 clade A *S. aureus* isolate, contains a phage approximately 38 kbp in size, with a gene encoding a teichoic acid biosynthesis protein. The sequence of the phage has 96 % nucleotide identity to regions of φNM1 in *S. aureus* strain Newman (Accession DQ530359.1), indicating that this is a bacteriophage of the *Siphoviridae* family. Other than the gene encoding a putative teichoic acid biosynthesis protein C, there were no virulence determinants identified on this phage, and the remaining genes encoded accessory phage functions including tail assembly proteins (Figure 4.1).

4.4.1.2 Arginine Catabolic Mobile element

The arginine catabolic mobile element (ACME) is a characteristic of some CA-MRSA clones, including the highly successful USA300 clone (Diep et al., 2006; 2008). ACME encodes proteins which may contribute to enhanced survival during community-associated infections, which has been attributed to the arginine deiminase system facilitating survival in acidic conditions such as those on human skin, and through the action of the SpeG protein, which inhibits host polyamine production during wound healing (Thurlow et al., 2013). All 7 UK isolates of the human CC97 clade B contained ACME linked to SCC*mec* type V (Ellington et al., 2008) (Figure 4.1).

4.4.1.3 Plasmids

The human clade A strains including the goat strain 42_2003 contained a plasmid encoding resistance to β-lactam antimicrobials, cadmium, and a putative bacteriocin

immunity protein (Figure 4.1). The plasmid is approximately 20 kbp in size and demonstrates 99 % nucleotide identity with the USA300_TCH959 plasmid pUSA300HOUMS (Accession number CP000732.1).

Strain 43_1980, the oldest strain from human clade B, also contains a 15 kbp class II plasmid with genes encoding cadmium and arsenic resistance and a bacteriocin immunity protein, but lacks the β-lactam resistance genes (Figure 4.1). The plasmid in strain 43_1980 has 99 % nucleotide identity with the homologous region of the 27 kbp *S. aureus* plasmid SAP019A (Accession number GQ900385.1). The plasmids in human clades A and B have limited sequence homology, indicating independent acquisition of plasmids from different origins (Figure 4.1).

Strain 45_2011 (Danish CC97 human clade A) contains a large conjugative plasmid of approximately 37 kbp in size that is unique to this strain, which has 99 % nucleotide identity to *S. aureus* plasmid pWBG745 (Accession GQ900389.1). In addition to genes encoding conjugation proteins such as the traE ATPase, a type IV secretion system (T4SS) virD4 domain was identified in the conjugative traG/traD family protein encoded on this plasmid. Previously conjugative systems have been described as a subfamily of T4SS within bacteria, and implicated in facilitating the spread of antimicrobial resistance and virulence genes among bacterial populations (Bhatty et al., 2013). Consistent with this, potential virulence genes located adjacent to the conjugation system in the plasmid of 45_2011 include genes encoding a secretory antigen SsaA-like protein and a Type IV secretory pathway adhesin AidA (Figure 4.1).

4.4.1.4 SCC*mec*

The presence of SCC*mec* was specific to human and pig CC97 strains, and absent from the bovine CC97 strains (For more detail see Section 3.4.4) (Figure 4.1).

4.4.2 Human CC97 *S. aureus* lack MGEs associated with bovine CC97 *S. aureus*

4.4.2.1. Livestock-associated SaPIs

Of the 20 CC97 *S. aureus* strains isolated from livestock, 8 strains including the pig strains were found to contain the SaPI encoding an allelic variant of von Willebrand factor binding protein (vWbp), which is distinct from the chromosomally-encoded *vwb* gene found in the majority of strains (Figure 4.1). The *vwb* gene contained on this SaPI (variants of which have been named as SaPIbov4, SaPIbov5, SaPIov2 and SaPIeq1) has been shown to confer the ability to coagulate ruminant plasma (Guinane et al., 2010; Viana et al., 2010) (Figure 4.1). Adjacent to the *vwb* gene there is *scn*, encoding a staphylococcal complement inhibitor, which has also been described previously (Guinane et al., 2010). In the CC97 *S. aureus*, the integration site is at the 3' end of the *gmps* gene encoding GMP synthase, which is the same integration site for SaPIbov1 in RF122 (Herron-Olson et al., 2007) and SaPIov1 in ED133 (Guinane et al., 2010). Comparison of representative SaPI sequences for the bovine and pig SaPIs indicate that both SaPIs look most similar in terms of nucleotide sequence and gene arrangement to SaPIbov5 (Figure 4.2B). The bovine CC97 SaPI shares 99 % nucleotide identity with 86 % of the length of SaPIbov5, while the pig CC97 SaPI shares 99 % nucleotide identity with 98 % of the length of SaPIbov5 (Figure 4.2B).

Of the 17 bovine CC97 *S. aureus* examined, 12 contained SaPIbov3 (Figure 4.1) which integrates at the same site as SaPIbov3 or SaPI1 in other *S. aureus* strains (Baba et al., 2002; Herron-Olson et al., 2007; Holden et al., 2004), adjacent to the groEL gene (Figure 4.2C). SaPIbov3 encodes a multidrug transporter (Herron-Olson et al., 2007).

In contrast, at the same integration site adjacent to the groEL gene, 15 of the 16 human clade A CC97 *S. aureus* strains contain a SaPI that is distinct from SaPIbov3 (Figure 4.2C). Examination of the closest homolog indicated that 64 % of the SaPI sequence demonstrates 98 % nucleotide identity to a SaPI from human strain 6850 (ST50) (Accession number CP006706) that integrates at the same site adjacent to the gene encoding groEL. In particular, in the region of the SaPI that is usually reported to contain virulence genes, there are 3 conserved genes sharing 100 % nucleotide identity (RSAU_001842-44). RSAU_001843 and RSAU_001844 encode hypothetical proteins, while RSAU_001842 shares 100 % nucleotide identity to an intracellular serine protease domain protein from *S. aureus* (Accession WP_001836981) (Figure 4.2C).

4.4.2.2 Bacteriophages

Bovine strains 40_2007, 5_2003, and 7_2003 contain a phage encoding the bi-component leukotoxin lukM/lukF'-PV which has been shown to have ruminant-specific leukotoxic activity (Barrio et al., 2006). None of the human or pig strains contained this phage (Figure 4.1).

In addition, sequence specific for a novel phage element was identified in 4 of the17 bovine strains and in both pig strains, but was not present in any of the human CC97 strains (Figure 4.1). This region was approximately 16.8 kbp in length, and encoded general phage proteins such as a capsid protein, holin and phage terminase (Figure 4. 1). Interrogation of this sequence against the Genbank database did not return any significant homologs.

4.4.2.3 Plasmids

Examination of plasmid content within the bovine CC97 strains revealed that 3 of the 17 strains contained plasmids (Figure 4.1). Bovine strains 37_1993_91 and 38_1993_91, both isolated in Ireland, contained a plasmid that has 99 % nucleotide identity to a 9 kbp region of *S. aureus* plasmid pWBG762 (Accession GQ900475). Genes of note on this plasmid encode a multidrug transporter integral membrane protein LmrP, a multidrug ABC transporter, in addition to metabolic enzymes involved in thiamine biosynthesis, a putative acyltransferase, and oligopeptidase (Figure 4.1, designated "bovine plasmid 1").

Bovine strain 36_1993_93, also isolated in Ireland, contained a conjugative plasmid specific to this strain, of which 10.9 kbp had 91 % nucleotide identity to regions of *S. aureus* strain 004-737X plasmid pSA737 (KC206006), and a 3 kbp region had 97 % nucleotide identity to *S. aureus* ED98 plasmid pAVX(CP001784). In addition to genes encoding conjugative proteins, there were 22 hypothetical proteins encoded and no known virulence determinants (Figure 4.1, designated "bovine plasmid 2").

4.4.2.4. SCC element

Bovine CC97 strains 9_1987 and 15_2006_75 have a novel SCC element that contains an arsenic resistance operon, and cassette chromosome recombinase genes ccrA and ccrB, but lacks the *mec* gene complex (See section 1.5.2.3) (Figure 4.1). The element is inserted at the same integration site as SCC*mec*, at the 3' end of the *orfX* gene, which is the known insertion site for other previously identified SCC elements (Chongtrakool et al., 2006; Holden et al., 2004; Luong et al., 2002).

4.4.3 Pig CC97 isolates share MGE with bovine and human CC97 isolates

Examination of the accessory genome of the pig strains revealed that they have several MGEs in common with human CC97 and several MGEs in common with bovine CC97, in addition to pig-specific MGE (Figure 4.1). For example, SCC*mec* type V was present in porcine and human clade B strains (Figure 4.1). A gene known as *czrC*, encoding cadmium zinc resistance protein C was found on the SCC*mec* element, which has been previously described on the SCC*mec* type Vc element of CC398 *S. aureus* (Price et al., 2012). In addition a plasmid containing the *tetK* gene was found in porcine and 3 human clade A isolates (Figure 4.1).

Porcine and bovine CC97 strains both contained the SaPI encoding vWbp, and the novel phage sequence that was identified in 4 of the bovine strains, which was also present in both pig CC97 strains (Figure 4.1). In addition, both pig strains have Tn552 encoding beta-lactamase resistance genes which is also found in bovine strain 12_1987 (Figure 4.1).

Several MGE were unique to the porcine strains, including a transposon containing the *tetM* gene, similar to that identified in LA-MRSA CC398 strains S0385 (Schijffelen et al., 2010) and 08BA02176 (Golding et al., 2012) (Figure 4.1).

A bacteriophage of the *Siphoviridae* family was identified that was specific to both pig strains and not contained in the CC97 bovine or human strains (Figure 4.1, pig-specific phage column). The CC97 pig bacteriophage was approximately 41 kbp in size, and had a 17.5 kbp region that demonstrated highest nucleotide identity (95 %) to prophage ϕSa1 of *S. aureus* strain Z172 (Accession number CP006838). Across the 41 kbp phage of the CC97 pig strains, genes encoding for accessory phage functions were found, with no known virulence determinants identified.

Figure 4.1. MGE distribution in CC97 *S. aureus* strains (See next page).
Bayesian phylogenetic tree (Section 3.4.3), with date and country shown and human clades A and B as labelled. MGE are grouped at the top of the diagram, with presence or absence indicated by a filled or empty box respectively. SCC*mec* type (type IV or V) and φSa3 IEC type (B or E) are shown. Columns are grouped according to MGE and colour-coded as follows; Red = SCC elements, Blue=SaPIs, Green=Phage, Purple=Plasmids and Magenta=transposons. Columns are numbered as follows: 1=SCC*mec*, 2=SCC, 3=ACME, 4=SaPI(vWbp), 5=SaPIbov3, 6=SaPI1, 7= φSa3, 8=lukM/lukF'-PV phage, 9=novel phage, 10=human-specific phage, 11=pig-specific phage, 12=*tetK* plasmid, 13=β-lactam plasmid, 14=heavy metal resistance plasmid, 15=conjugative plasmid, 16=bovine plasmid 1, 17=bovine plasmid 2, 18=*tetM* transposon, 19=Tn552.

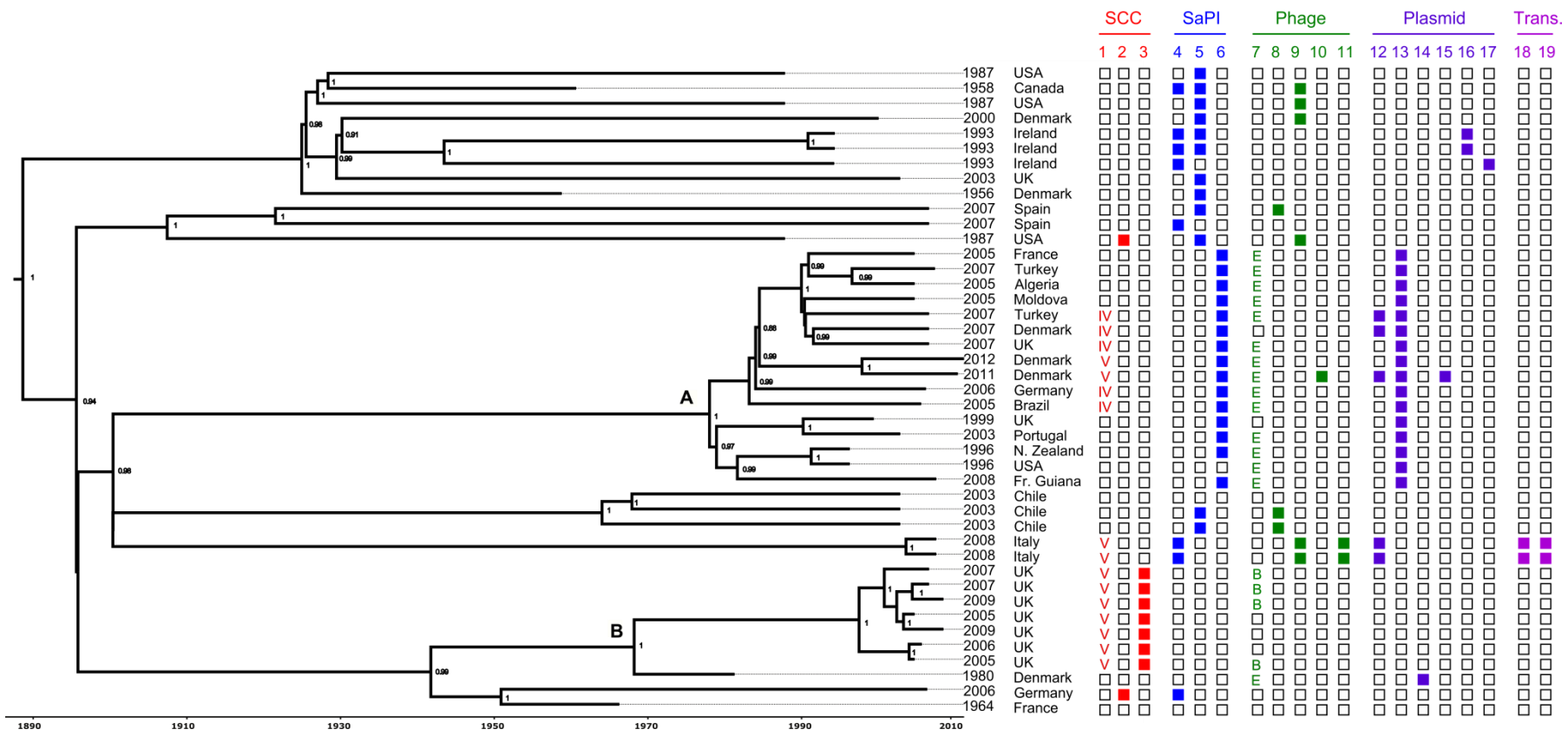
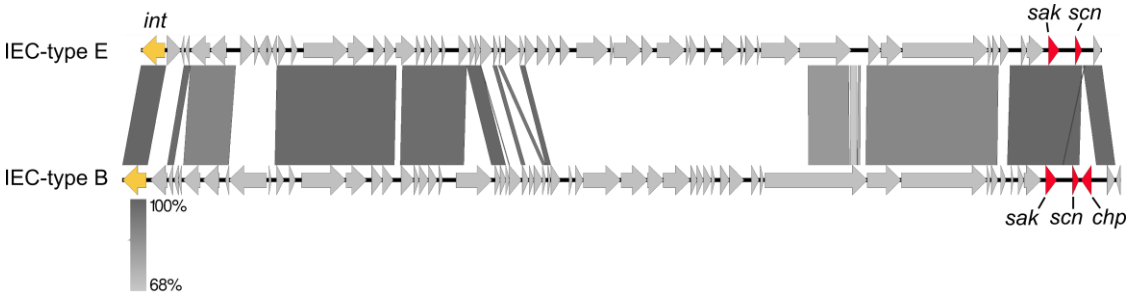
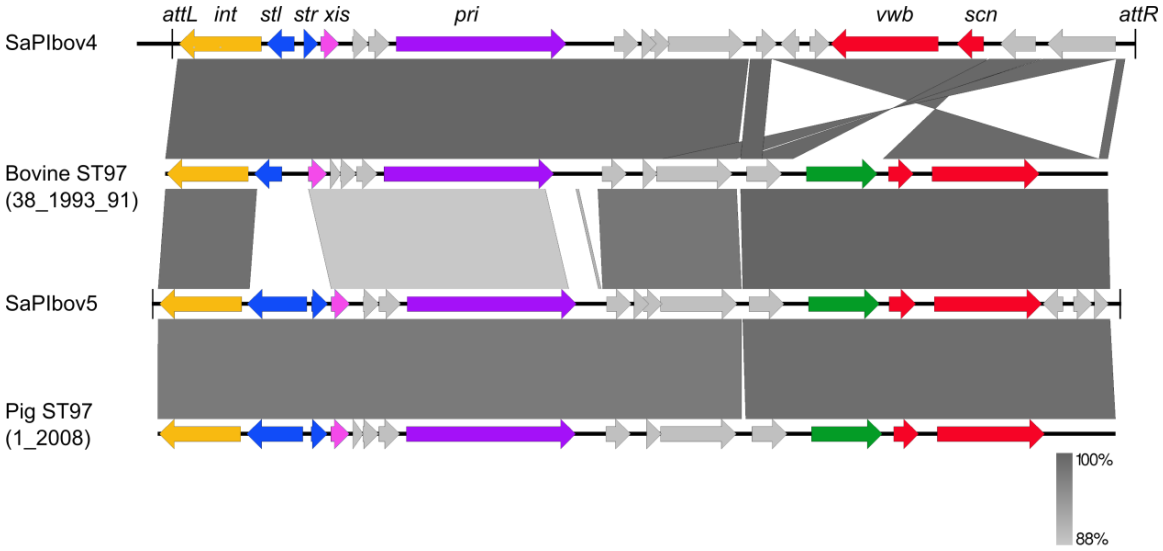

SCC
SaPI
Phage
Plasmid
Trans.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
1987 USA
1958 Canada
1987 USA
2000 Denmark
1993 Ireland
1993 Ireland
1993 Ireland
2003 UK
1956 Denmark
2007 Spain
2007 Spain
1987 USA
2005 France
2007 Turkey
2005 Algeria
2005 Moldova
2007 Turkey
2007 Denmark
2007 UK
2012 Denmark
2011 Denmark
2006 Germany
2005 Brazil
1999 UK
2003 Portugal
1996 N. Zealand
1996 USA
2008 Fr. Guiana
2003 Chile
2003 Chile
2003 Chile
2008 Italy
2008 Italy
2007 UK
2007 UK
2009 UK
2005 UK
2009 UK
2006 UK
2005 UK
1980 Denmark
2006 Germany
1964 France
A
B
1890
1910
1930
1950
1970
1990
2010

Figure 4.2 (See next page). Comparison of the genetic structure of MGE identified among CC97 strains in A) φSa3, B) $SaPI_{vWbp}$. and C) SaPIbov3 and SaPI1. Strain identity is indicated on the left of each diagram. The grayscale refers to pairwise percentage nucleotide identity. Genes are colour-coded as follows: integrase = yellow, IEC and virulence genes = red, SaPI transcriptional regulators= blue, *xis* gene= pink, SaPI replication module=purple, packaging module =green, hypothetical proteins=grey. *attL* and *attR* indicate left and right SaPI attachment sites.

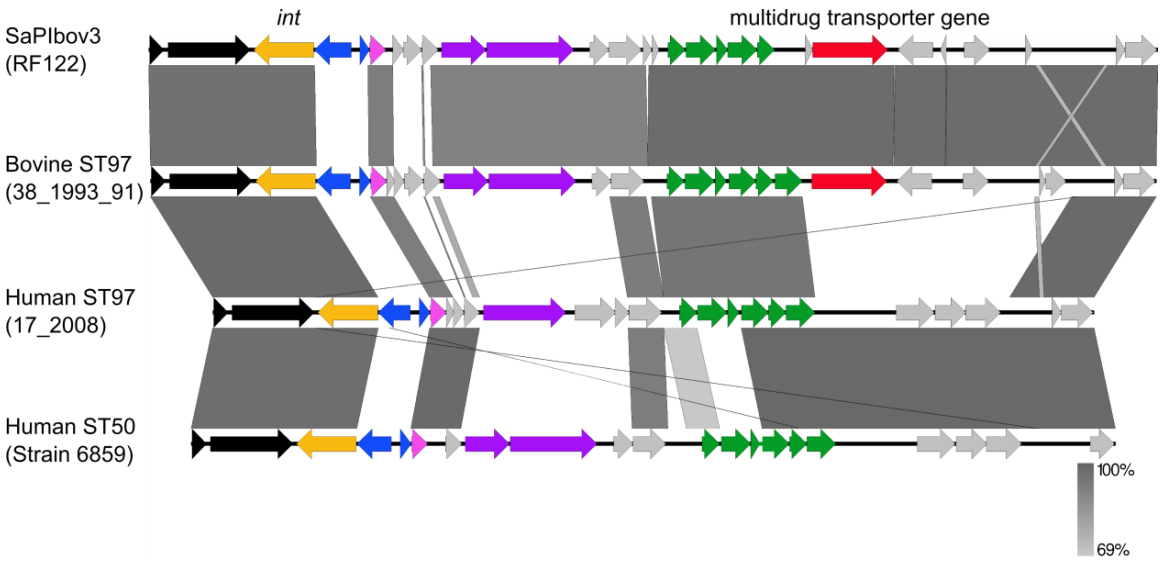
A)


int
sak scn
IEC-type E
IEC-type B
100%
68%
sak scn chp

B)


attL int stl str xis
pri
vwb
scn
attR
SaPIbov4
Bovine ST97
(38_1993_91)
SaPIbov5
Pig ST97
(1_2008)
100%
88%

C)


int
multidrug transporter gene
SaPIbov3
(RF122)
Bovine ST97
(38_1993_91)
Human ST97
(17_2008)
Human ST50
(Strain 6859)
100%
69%

4.4.4 Examination of genetic polymorphisms that may be associated with host adaptation in human *S. aureus* CC97

4.4.4.1 Nonsynonymous SNPs associated with host species

The phylogenetic analysis of the CC97 *S. aureus* lineage revealed at least 2 independent livestock-to-human host jumps (Section 3.4.3). Given that at least one of the human clades appears to have undergone intercontinental dissemination, this indicates that the strains have adapted successfully to humans and are capable of human to human transmission and global spread. As part of the current study, core SNPs and small insertions and deletions (indels) in CC97 *S. aureus* strains isolated from bovine, human and pigs were examined, in order to identify polymorphisms that might reflect host-adaptive changes occurring in the human strains following the host jumps. This included examination of nonsynonymous SNPS and polymorphisms resulting in pseudogenes, such as those introducing premature stop codons or resulting in a frameshift. In addition, any pseudogenes identified in livestock were examined for reversion to function in the human strains.

Examination of the distribution of 9422 phylogenetically informative SNPs among the CC97 *S. aureus* lineage indicated that there were 4035 nonsynonymous SNPs, 3048 synonymous SNPs, 2199 intergenic SNPs and 140 SNPs attributed to other effects, such as the introduction of stop codons, or loss of a start or stop codon. Examination of clade-associated SNPs indicated that in human clade A there were 88 nonsynonymous SNPs situated within 84 genes that were specific to this clade. Of these, 11 SNPs were contained in genes which did not exhibit allelic variation in any of the CC97 livestock strains, suggesting that these polymorphisms could have a host-specific role in these candidate genes (Table 4.1). This included the *fruB* gene, encoding fructose 1-phosphate kinase, which is involved in fructose and mannose metabolism, and 10 hypothetical proteins (Table 4.1).

In human clade B, 55 nonsynonymous SNPs situated within 55 genes were identified as clade-specific. Of these, 7 nonsynonymous SNPs were in genes that had no other livestock-associated polymorphisms, suggesting a possible host-specific role. This included the gene *murC*, encoding UDP-N-acetylmuramate-L-alanine ligase, which is involved in peptidoglycan biosynthesis, and gene *ampS* encoding an aminopeptidase enzyme involved in amino acid transport and metabolism, in addition to 5 hypothetical proteins (Table 4.2). There were no SNPs or genes identified in the candidate shortlists that were shared between the 2 clades, indicating a lack of convergent evolution. It should be pointed out that it is entirely feasible that the clade-specific polymorphisms are entirely neutral and that they arose by random mutation and genetic drift.

Table 4.1. **SNPs specific for human clade A *S. aureus* CC97 strains.**

Coordinate	Reference base	Clade base	Snp/Indel	Gene (MW2 reference)	Gene name/product	Snp effect	Codon change	Amino acid change
503032	G	A	SNP	MW0443	Hypothetical protein	Nonsynonymous	Gta/Ata	V/I
588980	C	T	SNP	MW0508	Hypothetical protein	Nonsynonymous	Cgt/Tgt	R/C
742777	A	C	SNP	MW0661	*fruB*	Nonsynonymous	gAt/gCt	D/A
797467	C	T	SNP	MW0710	Hypothetical protein	Nonsynonymous	Gat/Aat	D/N
996288	C	G	SNP	MW0902	Hypothetical protein	Nonsynonymous	gCg/gGg	A/G
1340066	C	T	SNP	MW1227	Hypothetical protein	Nonsynonymous	gCc/gTc	A/V
1794538	C	A	SNP	MW1657	Hypothetical protein	Nonsynonymous	Gtt/Ttt	V/F
1834106	C	T	SNP	MW1687	Hypothetical protein	Nonsynonymous	Gac/Aac	D/N
1931749	A	G	SNP	MW1777	Hypothetical protein	Nonsynonymous	gTc/gCc	V/A
2124735	C	A	SNP	MW1975	Hypothetical protein	Nonsynonymous	ttG/ttT	L/F
2645804	A	C	SNP	MW2480	Hypothetical protein	Nonsynonymous	cTa/cGa	L/R

Table 4.2. SNPs specific for human clade B *S. aureus* CC97 strains

Coordinate	Reference base	Clade base	Snp/Indel	Gene (MW2 reference)	Gene name/product	Snp effect	Codon change	Amino acid change
953226	T	C	SNP	MW0862	Hypothetical protein	Nonsynonymous	Ttt/Ctt	F/L
1009466	C	A	SNP	MW0915	Hypothetical protein	Nonsynonymous	Caa/Aaa	Q/K
1475135	C	T	SNP	MW1327	Hypothetical protein	Nonsynonymous	atG/atA	M/I
1828274	G	A	SNP	MW1683	*murC*	Nonsynonymous	gCg/gTg	A/V
1984007	G	A	SNP	MW1819	*amps*	Nonsynonymous	gCt/gTt	A/V
2387041	C	T	SNP	MW2241	Hypothetical protein	Nonsynonymous	Gaa/Aaa	E/K
2618541	T	A	SNP	MW2458	Hypothetical protein	Nonsynonymous	aAa/aTa	K/I

4.4.4.2 Pseudogenes identified within *S. aureus* CC97

Genetic variation resulting in predicted pseudogenes were examined in the CC97 strains. Strain-dependent variation in the number and distribution of pseudogenes was observed, ranging from 35 pseudogenes in strain 31_2007 to 66 predicted pseudogenes in strain 6_2003. Of these, nonsense mutations leading to predicted truncated genes encoding for 3 hypothetical proteins were identified as lineage-specific to all CC97 *S. aureus* strains. In addition, there were 2 SNPs in each human clade that had resulted in a premature stop codon (Table 4.3). In human clade A, the 2 pseudogenes encode hypothetical proteins, while in human clade B there is one pseudogene for a hypothetical protein, and one pseudogene encoding a glycerate dehydrogenase enzyme involved in coenzyme metabolism for energy production (Table 4.3). There were no indels resulting in frameshift mutations that were specific to either human clade. In addition, there was no evidence for reversion to function of any bovine pseudogenes in the human strains.

Table 4.3. **Pseudogenes specific to CC97 human clades A and B**

Human Clade	Coordinate	Reference base	Clade base	Snp/Indel	Gene (MW2 reference)	Gene name/product	Snp effect	Codon change	Amino acid change
A	1196147	T	C	SNP	MW1096	Hypothetical protein	Stop gained	tCa/tAa	S/*
A	2662541	G	A	SNP	MW2496	Hypothetical protein	Stop gained	Caa/Taa	Q/*
B	749262	G	T	SNP	MW0665	Hypothetical protein	Stop gained	Gga/Tga	E/*
B	897268	G	T	SNP	MW0812	Glycerate dehydrogenase	Stop gained	Gga/Tga	G/*

4.5. Discussion

Genomic changes associated with *S. aureus* host switches have been reported previously, including the acquisition of MGEs, allelic diversification of core genes and the formation of pseudogenes, all of which may alter the phenotype of the bacterium to facilitate survival and persistence in the new host (Guinane et al., 2010; Herron-Olson et al., 2007; Lowder et al., 2009). To date, most of our knowledge regarding *S. aureus* host adaptation has arisen from the analysis of adaptation to livestock hosts, following a transition from a human-associated ancestor (Guinane et al., 2010; Herron-Olson et al., 2007; Lowder et al., 2009).

However, the genetic events relating to *S. aureus* adaptation to the human host following a host jump has not yet been investigated. Since CC97 *S. aureus* has undergone at least 2 recent independent livestock-to-human host jumps, this provided a platform by which host-adaptation to the human host could be evaluated. Identification of the genetic determinants that are critical to adaptation to the human host might provide novel targets for intervention measures and control of *S. aureus* infections.

In the current study, examination of the variation in MGE content between CC97 *S. aureus* from different human and animal hosts revealed important differences that would be predicted to alter the ability of the human strains to evade human innate immune defences and persist in an environment with antimicrobial selective pressures. For example, the majority of human CC97 strains in the current study have acquired the β-toxin converting phage following the host jump from cows into humans, with at least 2 independent acquisitions of this phage, suggesting that the presence of this phage confers a fitness advantage in the human host. The β-toxin converting phage (ϕSa3) contains various complements of genes collectively termed the "Immune evasion complex" (IEC), the products of which have been shown to mediate evasion of human innate immune defences (Rooijakkers et al., 2005; 2005b;

Wamel et al., 2006). The findings of the current study are consistent with previous reports that have examined the prevalence of the ϕSa3 phage in human and animal *S. aureus* strains, which is strongly associated with the human host, and has been identified in over 90 % of human strains and a minority of livestock-associated strains (Monecke et al., 2007; Sung, Lloyd and Lindsay, 2008; Wamel et al., 2006).

Of note, goat strain 42_2003 was found to have the β-converting phage and the complement of MGE overall, in addition to the phylogenetic position of this strain, is consistent with a human origin for strain 42_2003. It is possible that this strain is a human contaminant or representative of a very recent goat-to-human transmission event. It is not known if this strain was isolated from a goat that was hand-milked, or an intensively-farmed goat from a large dairy farm. Further phylogenetic analysis of additional *S. aureus* CC97 goat strains may help clarify if zoonotic transmission or host jump events are contributing to the isolation of CC97 from goats.

Both the human clade A strains and the oldest human clade B strain contained different plasmids containing heavy metal resistance genes. Plasmids encoding heavy metal resistance have been identified in *S. aureus* strains of both human and animal origin and may not mediate host adaptation (Shearer et al., 2011). However, the lack of heavy metal resistance genes identified among the bovine isolates likely reflects the different ecological niches occupied by these strains. In addition to acquiring MGEs known to be important for human pathogenesis, the comparative genomic analysis carried out also indicated that the human CC97 strains have lost MGE that are known to be important in pathogenesis of *S. aureus* in the ruminant host. This included the prophage which encodes lukM/lukF'-PV, which has enhanced cytotoxic activity for ruminant leukocytes (Barrio et al., 2006), and was absent in both pig and human strains. The presence of a SaPI encoding a variant vWbp that specifically coagulates ruminant plasma would not likely be of benefit in the human host (Guinane et al., 2010; Viana et al., 2010). The lack of bovine MGE identified in human CC97 strains suggests that they do not confer a fitness advantage to the

strains in their new host environment. The human clade A strains contain a SaPI integrated at the same site as SaPIbov3, which has likely been acquired following the host jump, and may encode products with a human host-specific role.

The current study findings indicate that each host-associated CC97 subclone contains unique phage content (Figure 4.1). In the case of the β-converting phage the virulence genes, insertion site and function are well characterised (Wamel et al., 2006). However in others, such as the novel phage identified in the bovine and pig CC97 strains, further characterisation of the phage sequence and flanking regions is required to identify the insertion site, gene content and integrase gene sequence homology, which will facilitate characterisation of the phage family (Goerke et al., 2009). The fact that the distribution of phage correlates with host species implies that they may contain genes encoding proteins that have a role relating to host adaptation.

Previous *S. aureus* genome sequencing of livestock *S. aureus* strains has identified the loss of function of genes such as *SpA*, which are known to be important in human disease pathogenesis (Herron-Olson et al., 2007; Lowder et al., 2009). In the current study, polymorphisms resulting in pseudogenes were compared between livestock and human CC97 strains. There were 2 pseudogenes each that were specific to either the ruminant or human host, implying that this is not a large feature of adaptation associated with the identified host switches that have occurred in the CC97 lineage. There was no evidence that any bovine pseudogenes had reverted to function following the host jump into humans, and no virulence genes that might have altered function as the result of a nonsynonymous SNP resulting in an amino acid change. For the 3 pseudogenes identified in all CC97 strains, it is possible that these may have been an original adaptation to the cow host, and have not reverted back to function following the host switch to humans. The method used in the current study is quite conservative, and involved narrowing down a large list of variants to the current shortlist of candidate mutations that may be related to host adaptation. It is feasible, however, that a single mutation can have a large phenotypic effect. For

example, a single nonsynonymous point mutation in the *fusA* gene can effect a phenotype of resistance by modifying the drug target EF-G (McLaws et al., 2011). Further study might include a genome-wide scan for genes under diversifying selection using GWAS bioinformatic methods such as those provided by plink (http://pngu.mgh.harvard.edu/~purcell/plink/) (Purcell et al., 2007).

Taken together, the findings of the current study suggest that the horizontal transfer of MGEs has played a central role in the host adaptation of CC97 *S. aureus* to different host species. The data imply that strains of animal origin can successfully adapt to humans and spread globally, highlighting the need for continued surveillance in livestock, and early identification of emerging clones in humans of public health importance. This also highlights an important intervention point, since the identification of the mechanisms underlying host adaptation provides potential targets for therapeutic or infection control measures.

5. Role of recombination in the evolution of a ruminant-associated single locus variant of ST97

5.1 Introduction

Previous population genetic studies have shown that natural *S. aureus* populations are highly clonal (Fitzgerald et al., 1997; Musser et al., 1990; Reinoso et al., 2008; Zadoks et al., 2000). Closely related *S. aureus* sequence types (ST) group together into clonal complexes, and diversify clonally through the accumulation of point mutations, with estimates that a given nucleotide site is 15 times more likely to change by mutation than by homologous recombination (Feil et al., 2003; Pérez-Losada et al., 2006). This is in contrast to other bacterial species such as *Neisseria meningitidis*, which undergo large amounts of homologous recombination with some estimates of recombination occurring 80 times more frequently than mutation (Feil et al., 1999).

However, homologous recombination does occur within *S. aureus* and has been observed to play a role in the evolution of certain *S. aureus* lineages. For example, the widespread multi-drug resistant hospital-associated lineage ST239 arose as result of a large-scale homologous recombination event between an ST30 donor and a ST8 recipient lineage (Robinson and Enright, 2004). Other human-associated *S. aureus* lineages identified as having undergone a large-scale homologous recombination event include ST34 and ST42 (Robinson and Enright, 2004). Recently, core genome recombination analysis of a wide collection of *S. aureus* clones has revealed that recombination in the core genome is more widespread than previously identified, with certain hotspots of recombination associated with MGE (Everitt et al., 2014). In addition, recombination has played an important role in forming mosaic allelic variation in important virulence genes such as the superantigen *selx* (Wilson et al., 2011), and fibronectin binding protein genes *fnbpA* and *fnbpB* (Burke et al., 2010; McCarthy and Lindsay, 2010).

Since all previous studies of *S. aureus* that have identified large-scale recombination have included human-adapted lineages, the role of large-scale recombination in the

evolution of animal-associated *S. aureus* lineages is currently unknown. The ancestral genotype of CC97 is predicted to be ST97, and previous analysis of the level of mutation and recombination based upon MLST sequences for ST97 and single locus variants reveals evidence of a reticulate population structure within CC97 according to splits-graph analysis, with a roughly equal chance of alleles estimated to change by point mutation and recombination, which is much higher than predicted for *S. aureus* in general (Smith et al., 2005). Of note, the ruminant-associated lineage ST71, although a single locus variant of ST97, differs at the *arcC* MLST locus, which congruence analysis of MLST alleles has shown is the allele most likely to undergo recombination (Feil et al., 2003). Of the 6 SLVs of ST97 that differ at the *arcC* locus (ST71, ST355, ST1072, ST1129, ST1179 and ST2379), ST71 appears to be a widespread bovine-associated clone, and has previously shown phenotypic differences to ST97 including the expression of different capsule types (Guinane et al., 2008; Smyth et al., 2009). Therefore in the current study, the genomes of 2 ST71 strains were compared to representative ST97 strains to determine the level of genetic diversity existing within CC97, and to examine the possible role of recombination in the diversification of this major ruminant-adapted lineage.

5.2 Aims

To determine how the ruminant-associated ST71 strain differs genetically to its predicted ancestral genotype ST97:

- Compare the phylogenetic relationship of ST71 strains to their predicted ancestral genotype ST97

- Perform recombination detection analysis to assess the extent to which recombination has affected the evolution of ST71

- Compare the genomic content between ST71 and ST97 strains to identify genetic differences which could influence phenotype

5.3 Materials and Methods

5.3.1 Bacterial isolates.

The ST71 *S. aureus* strains used in the current study were both from cases of bovine mastitis. Strain RF103 was isolated from a case of bovine mastitis in Ireland in 1993 (Fitzgerald et al., 1997), and strain CO1122 was isolated in the UK in 2003 (Sung, Lloyd and Lindsay, 2008). For comparative genomic analysis, these strains were compared to representative bovine mastitis-associated CC97 strains (Table 5.1). Additional comparisons of protein coding sequences were made to a selection of diverse *S. aureus* isolates representing a range of major human and animal genotypes for which genome sequences were available in public databases (Table 5.1).

Table 5.1. *S. aureus* strains used in this study

Strain ID	Country (Region)	Host	Sequence type (MLST)	Clonal complex	Accession Number (If available)	Genome sequence reference
CO1122	UK	Bovine	71	97	N/A	(Sung, Lloyd and Lindsay, 2008)
RF103	Ireland	Bovine	71	97	N/A	(Fitzgerald et al., 1997)
38_1993_91	Ireland	Bovine	97	97	ERS212283	This study
29_1964	France	Bovine	97	97	ERS212275	This study
12_1987	USA	Bovine	97	97	ERS212258	This study
8_2003	UK	Bovine	97	97	ERS212255	This study
6_2003	Chile	Bovine	97	97	ERS212253	This study
17_2008	French Guiana	Human	97	97	ERS212263	This study
24_2009	UK	Human	97	97	ERS212270	This study
28_2005_91	Brazil	Human	97	97	ERS212274	This study
30_2005	Moldova	Human	97	97	ERS212276	This study
46_2012	Denmark	Human	97	97	ERS249846	This study
1_2008	Italy	Pig	97	97	ERS212248	This study
40_2007	Spain	Bovine	352	97	ERS212285	This study
9_1987	USA	Bovine	124	97	ERS212256	This study

Strain ID	Country (Region)	Host	Sequence type (MLST)	Clonal complex	Accession Number (If available)	Genome sequence reference
Newbould305	Canada	Bovine	115	97	AKYW01000000	(Bouchard et al., 2012)
ED133	France	Ovine	133	133	NC_017337	(Guinane et al., 2010)
LGA251	UK	Bovine	425	425	NC_017349	(García-Álvarez et al., 2011)
RF122	Ireland	Bovine	151	151	NC_007622	(Herron-Olson et al., 2007)
ERR144792	Denmark	Bovine	130	130	ERR144792	(Harrison et al., 2013)
11819-97	Denmark	Human	80	80	NC_017351	(Stegger et al., 2012)
HO50960412	UK	Human	22	22	NC_017763	(Köser et al., 2012)
N315	Japan	Human	5	5	NC_002745	(Kuroda et al., 2001)
COL	UK	Human	8	8	NC_002951	(Gill et al., 2005)
04-02981	Germany	Human	225	225	NC_017340	(Nübel et al., 2010)
USA300 TCH1516	USA	Human	8	8	NC_010079	(Highlander et al., 2007)
MW2	USA	Human	1	1	NC_003923	(Baba et al., 2002)
M013	Taiwan	Human	59	59	NC_016928	(Huang et al., 2012)
JKD6159	Australia	Human	93	93	NC_017338	(Chua et al., 2010)
S0385	Netherlands	Human	398	398	NC_017333	(Schijffelen et al., 2010)

Strain ID	Country (Region)	Host	Sequence type (MLST)	Clonal complex	Accession Number (If available)	Genome sequence reference
JKD6008	New Zealand	Human	239	239	NC_017341	(Howden et al., 2010)
MN8	USA	Human	30	30	NZ_CM000952	Shotgun sequence unpublished
MRSA252	UK	Human	36	30	NC_002952	(Holden et al., 2004)
CA-347	USA	Human	45	45	NC_021554	(Stegger et al., 2013)
MSHR1132	Australia	Human	75	75	NC_016941	(Holt et al., 2011)

5.3.2 Genome sequencing, mapping assembly, and *de novo* genome assembly.

Paired end Illumina sequencing was carried out on a Genome Analyser IIx, following standard Illumina protocols. Nucleotide distribution and read quality was assessed using FastQC v0.10.0 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and filtered for quality using the suite of tools available within FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) (Table 2.2). Reads were aligned against the reference genome MW2 (NC_003923), a triple locus variant of ST97 with variants detected and consensus sequences called as described in section 3.3.2 (Table 2.2). As previously, core genome was defined as sites shared by all strains within the mapping analysis. The effect of variants was determined using SnpEff v3.0 (Cingolani et al., 2012) (Table 2.2). Pairwise analysis of SNP density was calculated in 10 000 bp windows across the length of the genome using a custom python script, which examined all nucleotide sites in the genome of each strain against the reference genome MW2 in windows of 10 000 bp, and, not including uncalled bases (designated as "n"), would count any base differing from the reference as a SNP, outputting those counts to a tab-delimited file. For the purposes of *de novo* assembly, low frequency erroneous reads were removed using Quake v0.3.4 (Kelley, Schatz and Salzberg, 2010) and *de novo* assemblies generated using Velvet v1.2.07 (Zerbino and Birney, 2008) using the VelvetOptimiser.pl script implemented in VelvetOptimiser v-2.1.7 (http://bioinformatics.net.au/software.velvetoptimiser.shtml) (Table 2.2).

5.3.3 Molecular typing and evolutionary analysis of protein coding sequences.

De novo assemblies and mapped consensus sequences for each strain were interrogated for the 7 MLST alleles and 7 *S. aureus* surface associated protein (*sas*)

genes (Table 5.2) using BLASTN (Altschul et al., 1990). MLST alleles for each strain were queried against the *S. aureus* MLST database (http://saureus.mlst.net/) and *sas* alleles against the Sas database respectively (kindly provided by Ashley Robinson, available on request) to determine MLST and *sas* type. Details on each *sas* allele are listed in Table 5.2. Where required the presence or absence of the *sasD* allele was verified by PCR (Section 2.3, Table 2.1). If required, to verify *sas* alleles, PCR and Sanger sequencing were carried out using methods described in section 2.3 and primers listed in Table 2.1.

An additional 5 CDS situated within and 5 CDS situated outwith the “SNP dense” region were chosen at random for phylogenetic analysis. The "SNP dense" region is the region of the ST71 genome demonstrating a high density of tightly clustered SNPs spanning the origin of replication relative to the ST97 strains (See Section 5.4.1). Nucleotide sequences for each CDS were extracted from *S. aureus* genome sequences using BLASTN (Altschul et al., 1990). Strains included in the analysis are as listed in Table 5.1. Nucleotide sequences were aligned using the ClustalW method and neighbour joining trees constructed using MEGA v4.0 (Tamura et al., 2007) (Table 2.2). Statistical support at each node was assessed using 1000 bootstrap replicates.

Table 5.2. Details of *sas* typing alleles

***Sas* allele**	**Allele size (bp)**	**Product**	**Reference (if function has been characterised)**
A	462	Serine rich adhesin for platelets (sraP)	Siboo et al, 2005
B	462	Homologous to fmtB (associated with methicillin resistance)	Komatsuawa et al, 2001
D	453	Putative cell wall surface anchor protein	-
E	450	Broad spectrum ligand protein, expressed under iron limiting conditions (isdA)	Clarke et al, 2004
F	439	Putative cell wall surface anchor protein	-
H	467	5' nucleotidase involved in adenosine synthesis (adsA)	Thammavongsa et al, 2009
I	467	Putative cell wall surface anchor protein	-

5.3.4 Whole genome phylogenetic analysis.

The ST71 consensus sequences were added to the CC97 consensus sequences of strains listed previously in Table 3.1, Section 3.3.1 and core genome redefined using the customised core genome python script described in Section 3.3.3. A maximum likelihood phylogenetic tree was constructed using RAxML-7.2.6 (Stamatakis, 2006), implementing a GTR model with gamma correction for rate heterogeneity and 1000 bootstrap replicates (Table 2.2). For comparison, a second analysis was performed on an adjusted alignment that was made by removing the putative large-scale recombinant region from the core genome alignment, and repeating the analysis using identical parameters. Network analysis was conducted by extracting SNPs from both alignments and using the SNP alignments as input into the program Splitstree v4.13.1 (Huson and Bryant, 2006) (Table 2.3). As a preliminary analysis for recombination, the phylogenetic network was analysed using the phi test, which was implemented within Splitstree.

5.3.5 Recombination detection.

In order to assess the level of recombination among predicted ruminant-associated lineages of *S. aureus,* an alignment of the study strains along with published ruminant-associated *S. aureus* genomes was created using the progressive Mauve algorithm implemented in Mauve v2.3.1 using default settings (Darling, Mau and Perna 2010) (Table 2.2). To assess the core genome among these strains, locally collinear blocks (LCBs) of at least 1000 bp in length were extracted from the XMFA file using the stripSubsetLCB script distributed with Mauve (http://gel.ahabs.wisc.edu/mauve/snapshots/2012/2012-06-07/linux-x64/). The LCB blocks were concatenated, ordered according to MW2 coordinates, and converted to a FASTA alignment file format. Recombination detection was performed using BratNextGen (http://www.helsinki.fi/bsg/software/BRAT-NextGen/) (Marttinen et al., 2012), setting the hyper parameter alpha to 1, with a cutoff value of 0.1 within

the PSA tree, conducting 40 iterations within the detecting recombination algorithm and performing 100 permutation runs when estimating significance (Table 2.2).

5.3.6 Comparative genome analysis of predicted recombinant region.

The population-based *de novo* assembly software, Cortex1.0.5.20 (Iqbal et al., 2012) was used to identify variation in gene content among ST71 and ST97 strains (Table 2.2). Cortex utilises coloured de Bruijn graphs to detect variant sequence among bacterial population datasets (Iqbal et al., 2012). The ST71 strains listed in Table 5.1 were defined as group 1, and the bovine ST97 strains as group 2. Variant sequences that were present in at least 1 strain of each group, but none of the strains in the comparison groups were identified, then filtered as representative of genotype only if that variant sequence was present in all strains (for example if a gene was present in all bovine ST97 but none of the ST71 strains).

Identified variant CDS were annotated using Prokka -1.5.2 (Prokka: Prokaryotic Genome Annotation System - http://vicbioinformatics.com/) and BLASTX against the nr Genbank database (Altschul et al., 1990) (Table 2.2). For each strain, *de novo* assemblies were aligned against *S. aureus* Newbould 305 (AKYW01000002) using the Mauve Contig Mover tool implemented in Mauve v.2.3.1 (Rissman et al., 2009) to identify which contigs align to the predicted recombinant region. Contig 002 of ST115 strain Newbould 305 (AKYW01000002) was chosen as a reference because this contig contains the whole region of interest and is a closely related single locus variant of ST97.

5.4 Results

5.4.1 A "SNP dense" region in the ST71 *S. aureus* strains indicates a distinct evolutionary origin

In order to determine the genetic relatedness of the ST71 strains to the predicted ancestral genotype, ST97, the genomes of 2 ST71 *S. aureus* strains were sequenced and added to the previously sequenced strains for phylogenetic analysis (See Table 5.1, Section 3.3.1). The core genome for all strains was redefined with the additional strains added to the alignment, and a maximum likelihood phylogeny was reconstructed using RAxML-7.2.6 (Stamatakis, 2006) (Figure 5.1A). The ST71 strains were found to cluster closely together, distant from the rest of the CC97 *S. aureus* strains, which included other single locus variants such as ST124 and ST115 (Figure 5.1A). Network analysis of the extracted SNP alignment using Splitstree revealed evidence of a reticulate structure at the junction of the bovine ST97 and ST71 strains (Figure 5.2A), with the phi test finding statistically significant evidence for recombination (p = 1.154E-7).

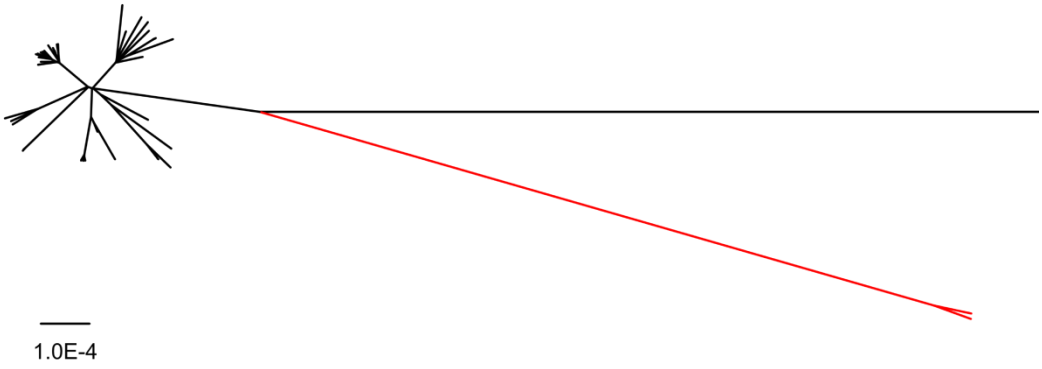
For a more in-depth analysis of the SNP distribution, a representative ST71 strain, RF103 and a representative bovine ST97 strain, 38_1993_91 were chosen and analysed in further detail, using the SNPs called against the MW2 reference. The core genome alignment for these 2 strains against MW2 consisted of 2472592 bp containing 4905 SNPs between RF103 (ST71) and 38_1993_91(38_1993_91). A region spanning the origin of replication containing a high density of tightly clustered SNPs was observed between the ST71 and ST97 strain, which was designated the "SNP dense" region (Figure 5.3 and Figure 5.4). Of the 4905 SNPs between the ST71 and ST97 strain, 4462 (91 %) were within the "SNP dense" region, while 443 SNPs (9 %) were outwith this region. Within the 4462 SNPs in the "SNP dense" region, 2766 (62 %) were synonymous, while 936 (21 %) SNPs were nonsynonymous SNPs (Table 5.3). In contrast, in the remainder of the genome, 229

of the 443 (51.6 %) SNPs were nonsynonymous, while the remaining 108 SNPs (24.5 %) were synonymous (Table 5.3).

Pairwise SNP density across the genome indicated that over the majority of the genome, RF103 (ST71) and 38_1993_91 (ST97) shared a similar pattern of SNP density (Figure 5.3A). However higher SNP numbers in RF103 were concentrated at the start and end of the genome, in the region spanning the origin of replication (Figure 5.3B). The individual SNP coordinates were plotted on a circular diagram, highlighting the dense clustering region of SNPs spanning the origin of replication (Figure 5.4). This "SNP dense" region of clustered SNPs started at coordinate 2689485, within a gene encoding fructose-1,6 bisphosphonate aldolase, spanned the origin of replication, and ended at coordinate 333931 within a gene encoding a hypothetical protein, encompassing approximately 15 % of the genome.

Taken together these data suggest that the "SNP dense" region in the ST71 strain may have a distinct evolutionary origin. This region was subsequently removed from the core genome alignment and the phylogenetic analysis repeated (Fig 5.1B). Where previously the ST71 strains had been situated as a separate clade distinct from the remainder of the CC97 strains (Fig 5.1A), they now clustered with the bovine CC97 strains (Fig 5.1B). Using Splitstree v4 (Huson and Bryant, 2006), the phi test was repeated after removing the SNP dense region, and did not find statistically significant evidence for recombination ($p = 0.838$) (Figure 5.2C).

A)



B)

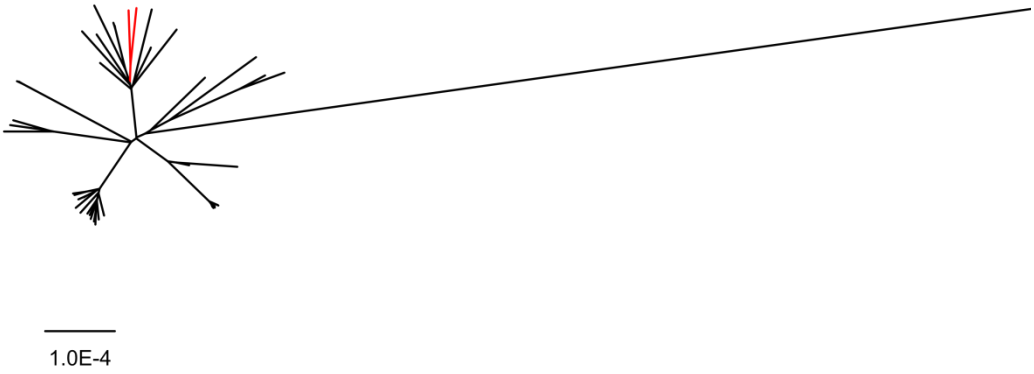
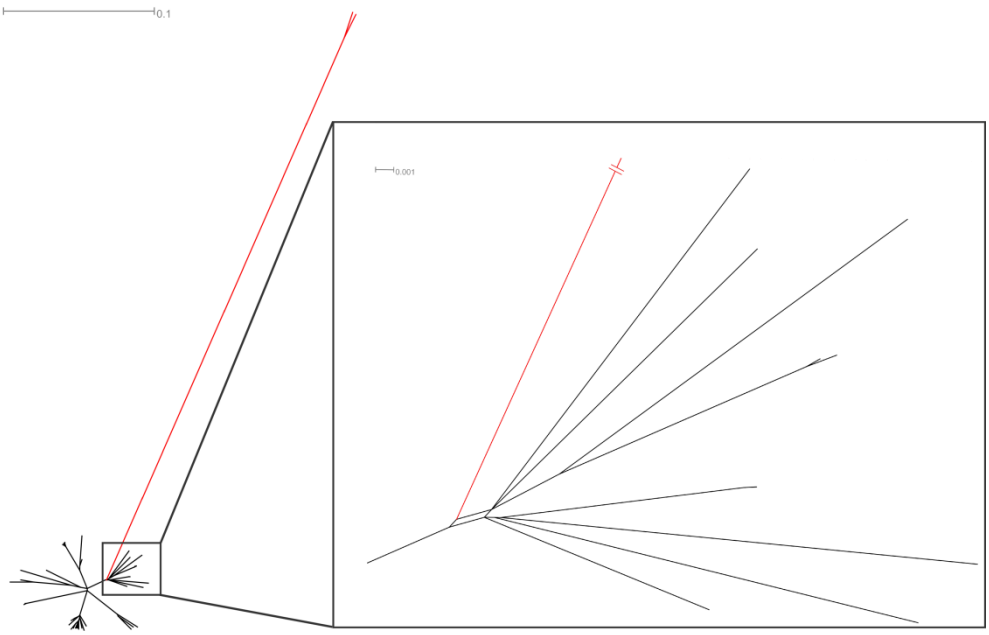


Figure 5.1. Radial maximum likelihood phylogeny of CC97 *S. aureus* strains including ST71. A) Maximum likelihood tree constructed from core genome alignment. B) Maximum likelihood tree constructed from core genome alignment with "SNPdense" region removed. Both radial trees are rooted against outgroup strains MW2 and 18_1997 (not shown), and the ST71 strains are highlighted in red. CC97 strains included are those listed in Table 3.1 Section 3.3.1 in addition to the ST71 strains. Scale is indicated for each tree in substitutions per site.

A)



B)

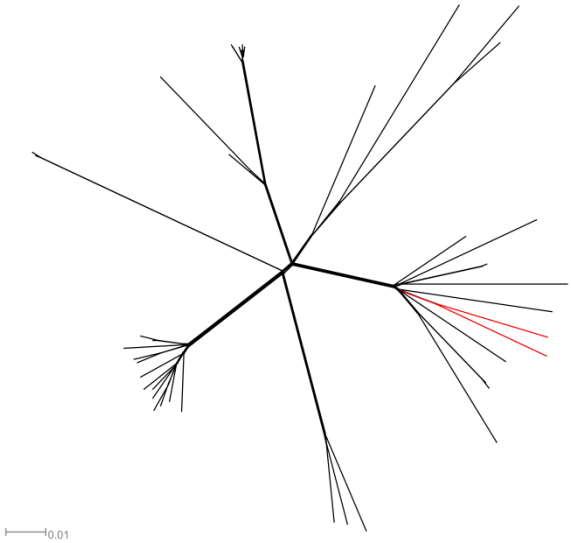
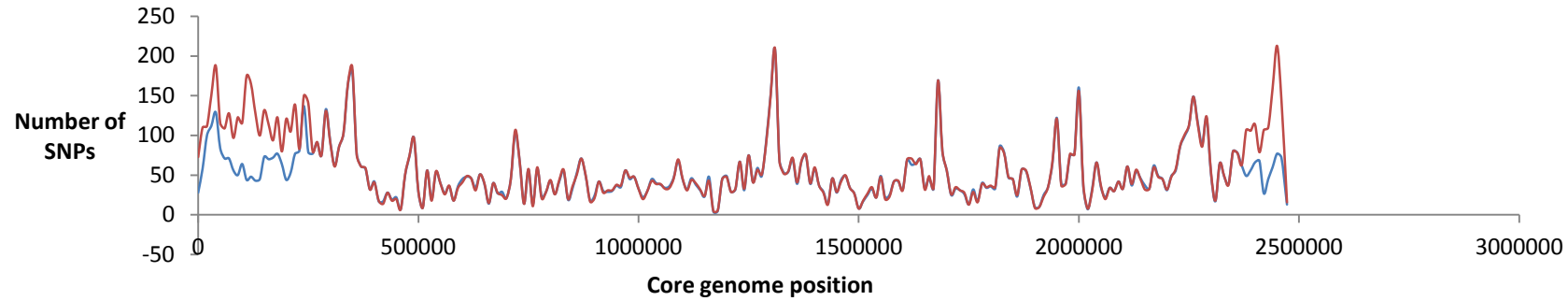


Figure 5.2. Splitstree analysis of core genome SNP alignment for CC97 and ST71 *S. aureus* A) including the "SNP dense" region and B) without the "SNP dense" region. Diagram A) indicates the whole tree and a zoomed in view of the nodes between the ST71 strains (red) and the largest clade of bovine CC97 strains (black). Strains included in both diagrams are ST71 (red) and other CC97 strains (Table 5.1 Section 3.3.1). Scale bars are as shown. Analysis conducted in Splitstree v4.0 (Huson and Bryant, 2006)

A)



B)

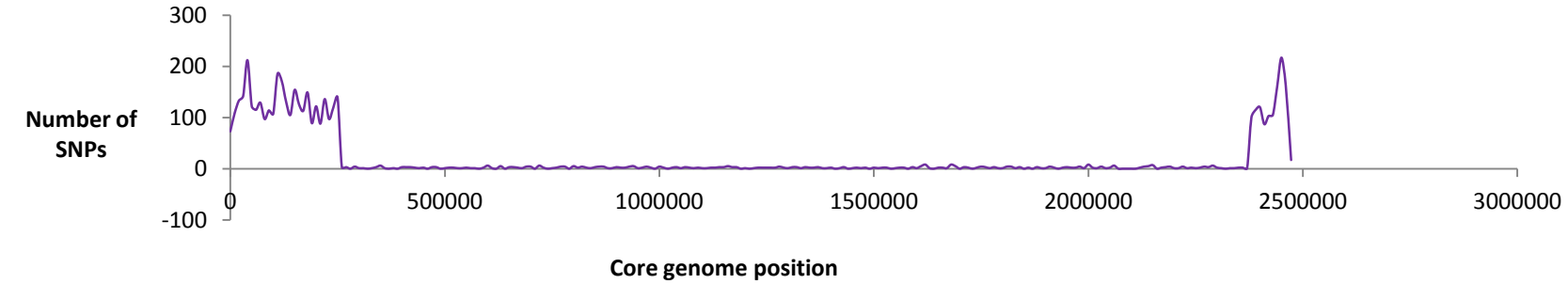


Figure 5.3. Pairwise strain comparisons of core genome SNP density over 10 000 bp windows. Graph A illustrates the number of SNPs per 10 000bp window for strain RF103 (ST71) compared to the reference strain MW2 (Red line, graph A) and strain 38_1993_91 (ST97) against MW2 (Blue line, graph A). Graph B compares core genome SNP density between strains RF103 and 38_1993_91 directly (purple line).

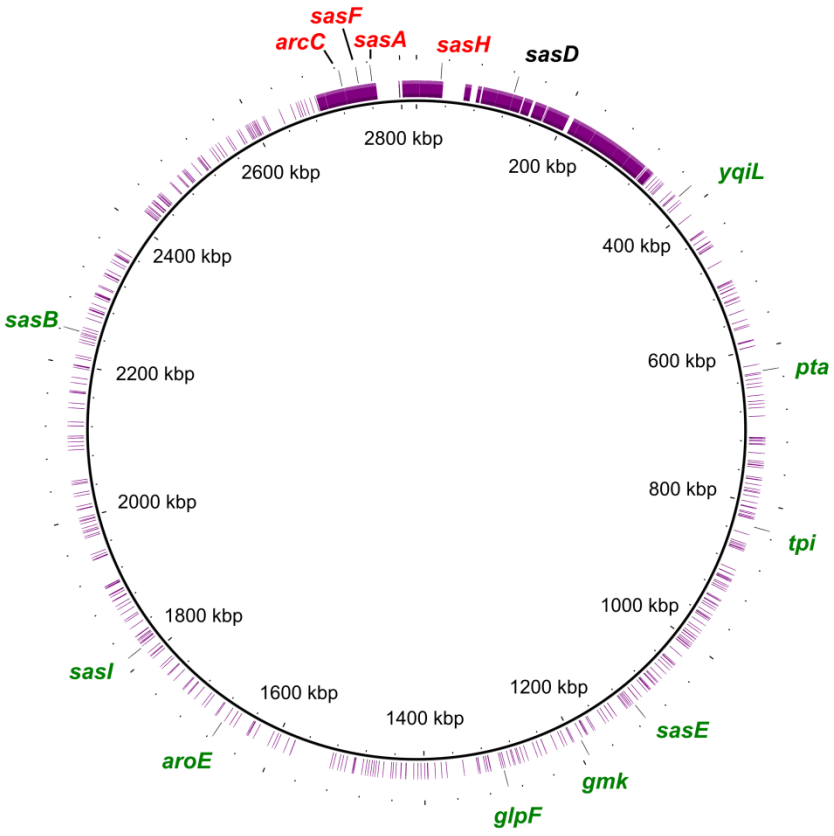


Figure 5.4. Circular diagram illustrating the genome distribution of SNPs between ST71 and ST97. Inner ring (purple) denotes coordinates of SNPs identified between consensus sequences for strains RF103 (ST71) and 38_1993_91 (ST97). Outer ring demonstrates genomic position of MLST and *sas* typing alleles, colour-coded according to the closest homolog that was identified for the ST71 strains (Red=Non-CC97 ruminant genotype, Green=ST97 genotype, Black=Absent). Diagram produced using BRIG v0.95 (Alikhan et al., 2011).

Table 5.3: SNP details comparing strain RF103 (ST71) and 38_1993_91 (ST97)

SNP effects	**Within SNP dense region**	**Outwith SNP dense region**
Nonsynonymous SNPs	936 (21 %)	229 (51.6 %)
Synonymous SNPs	2766 (62 %)	108 (24.5 %)
Intergenic	749 (16.8 %)	99 (22.3 %)
Other *	1 x StL (38_1993_91) 5 x SG (RF103) 3 x SG (38_1993_91) 1 x SpL (38_1993_91) 1 x SpL (RF103) (0.2 %)	1 x StL (38_1993_91) 3 x SG (RF103) 3 x SG (38_1993_91) (1.6 %)
TOTAL SNPs	**4462**	**443**

*****Other effect (the strain the effect is identified as follows in parentheses): Loss of a start codon (StL), gain of a stop codon (SG), loss of a stop codon (SpL). SNP effects were determined using SnpEff v3.0 (Cingolani et al., 2012).

5.4.2 Mosaic genome content in ST71 *S. aureus* strains indicates recombination with non-CC97 bovine-adapted lineages

In addition to MLST, an extended molecular typing scheme employing the *S. aureus* surface-associate proteins (*sas* typing) has also been used previously to type both human and animal strains of *S. aureus* (Robinson and Enright, 2003; Smyth et al., 2009). Sas typing involves comparison of the nucleotide sequence for regions of 7 *sas* genes (*sasA, sasB, sasD, sasE, sasF, sasH* and *sasI),* which encode surface-expressed proteins containing the LPXTG motif, involved in host-pathogen interaction (Clarke, Wiltshire and Foster, 2004; Siboo et al., 2005; Thammavongsa et al., 2009). For the purposes of the current study they were selected due to their distribution around the whole genome, with the genes *sasA, sasD, sasF* and *sasH* residing within the predicted recombinant region, and *sasB, sasE* and *sasI* in the non-recombinant region of the genome (See Figure 5.4 outer ring on circular diagram for genomic positions of MLST and *Sas* alleles).

Here, *sas* typing was conducted *in silico*, interrogating a database of known alleles (kindly provided by A. Robinson). The *sas* alleles for the ST71 strains were compared to *sas* alleles in a panel of CC97 strains, including ST97 strains from cows, pigs and humans, along with other single locus variants of CC97 (ST115, ST124 and ST352), and to non-CC97 ruminant genomes with publically available genome sequences (Table 5.1).

In the predicted "SNP dense" region, the ST71 strains contained *sas* alleles that were identical or demonstrated closest homology to a number of distinct non-CC97 bovine-adapted genotypes, although the clonal origin of the closest homolog varied depending on which gene was examined. (Table 5.4, Appendix 1 Figure S.1 A-F). For example, in addition to sharing the MLST *arcC*18 allele, both ST71 strains share the *sasA*13 allele with LGA251, an ST425 bovine strain (Table 5.4). *SasA*13 is

phylogenetically distinct from the *sasA*2 allele found in the remaining CC97 strains, as indicated on a neighbour joining phylogenetic tree (Appendix 1 Figure S.1A). In addition, neither of the ST71 strains contained *sasD*, which is also absent from several non-CC97 strains including ED133 (ST133) and LGA251 (ST425) (Table 5.4, Figure 5.5).

For the *sasF* allele, strain CO1122 (ST71) has *sasF*19, and strain RF103 has a novel *sasF* allele differing by 1 SNP from *sasF19*, which have closest homology to ovine CC133 strain ED133 (Appendix 1 Figure S.1D). The non-ST71 CC97 animal strains have distinct *sasF* alleles of *sasF*27 and *sasF*32 for animal and human CC97 strains respectively, which differ by 1 nonsynonymous SNP (Appendix 1 Figure S.1D). In addition, both ST71 strains share the *sasH15* allele with strain ED133 (ST133) (Appendices Figure 5.1E).

In contrast for the genes situated in the non "SNP dense" region of the genome, all alleles (*sasB*10, *sasE*8 and *sasI*10) were identical to the predicted ancestral lineage ST97 (Figure 5.4, Table 5.4), and were phylogenetically distinct from any of the alleles carried by the non-CC97 genotypes (Appendix 1 Figure S.1B,C,F).

To further investigate the potential evolutionary origins for the "SNP dense" region within the ST71 strains, the sequences for CDS situated within or outwith the "SNP dense" region were examined in more detail using phylogenetic trees constructed from DNA sequence alignments, comparing ruminant-associated strains of distinct genotypes (Figure 5.6). This included genes MW2524 and MW2526 flanking one side of the "SNP dense" region (Figure 5.6 A,B), and genes MW0286 and MW0288 flanking the other side (Figure 5.6 C,D). For the 2 CDS within the "SNP dense" region (MW2526 and MW0286), the sequences for the ST71 strains clustered most closely with sequences from non ST97 bovine genotypes (Figure 5.6B, 5.6C). The nucleotide sequence for MW2526 for the ST71 strains had 99 % nucleotide identity

to the closest homolog from the small ruminant CC133 strain, ED133 (Figure 5.6B). The closest homolog to the ST71 strains for MW0286 was from the bovine ST425 strain LGA251, showing 100 % nucleotide identity (Figure 5.6C). In contrast, for the 2 CDS outwith the "SNP dense" region (MW2524 and MW2088), the ST71 strains clustered most closely to the alleles from the ancestral genotype ST97 (Figure 5.6A, 5.6D). Further analysis using an additional 5 CDS within and 5 CDS outwith the "SNP dense" region were chosen at random and analysed using the same methods, confirming the finding that the region spanning the origin of replication has an evolutionary origin which is distinct to CC97 (Appendix 1 Figures S.2, S.3).

Table 5.4. Sas typing results for ST71, ST97 and non-CC97 ruminant *S. aureus* strains

Isolate	*sasA**	*sasB*	*sasD*	*sasE*	*sasF*	*sasH*	*sasI*	*Sas* type
CO1122 (Bovine ST71)	13	10	Absent	8	19	15	10	NT†
RF103 (Bovine ST71)	13	10	Absent	8	Novel (19)^	15	10	NT
38_1993_91 (Bovine ST97)	2	10	9	8	27	19	10	5
29_1964 (Bovine ST97)	2	10	9	8	27	19	10	5
12_1987 (Bovine ST97)	2	10	9	8	27	19	10	5
8_2003 (Bovine ST97)	2	10	9	8	27	19	10	5
6_2003 (Bovine ST97)	2	10	9	8	27	19	10	5
17_2008 (Human ST97)	2	10	9	8	32	19	10	Novel
24_2009 (Human ST97)	2	10	9	8	32	19	10	Novel
28_2005_91 (Human ST97)	2	10	9	8	32	19	10	Novel

Isolate	*sasA**	*sasB*	*sasD*	*sasE*	*sasF*	*sasH*	*sasI*	*Sas* type
30_2005 (Human ST97)	2	10	9	8	32	19	10	Novel
46_2012 (Human ST97)	2	10	9	8	32	19	10	Novel
1_2008 (Pig ST97)	2	10	9	Novel (8)	27	Novel (19)	10	NT
40_2007 (Bovine ST352)	2	10	22 (9)	8	27	19	Novel (10)	NT
9_1987 (Bovine ST124)	2	10	9	8	27	19	10	NT
Newbould305 (Bovine ST115)	32	10	9	8	27	19	10	12
ED133 (Ovine ST133)	23	29	Absent	27	Novel (34)	15	30	NT
LGA251 (Bovine ST425)	13	27	Absent	27	Novel (33)	Novel (16)	Novel (30)	NT
RF122 (Bovine ST151)	33	Absent	23	29	35	37	32	NT
ERR144792 (Bovine ST130)	26	27	24	26	33	36	28	NT

*Alleles highlighted in red are those situated in the "SNP dense" region of the ST71 genomes (See Figure 5.3 illustrating genomic coordinates for each allele). ^ For novel alleles, the closest homolog is shown in brackets, in all cases differing by 1 SNP. †NT=Non-typable. Due to one or more missing alleles, the *sas* type could not be determined. The *sas* type was deemed novel if all alleles were present, but not identified by the querying the database.

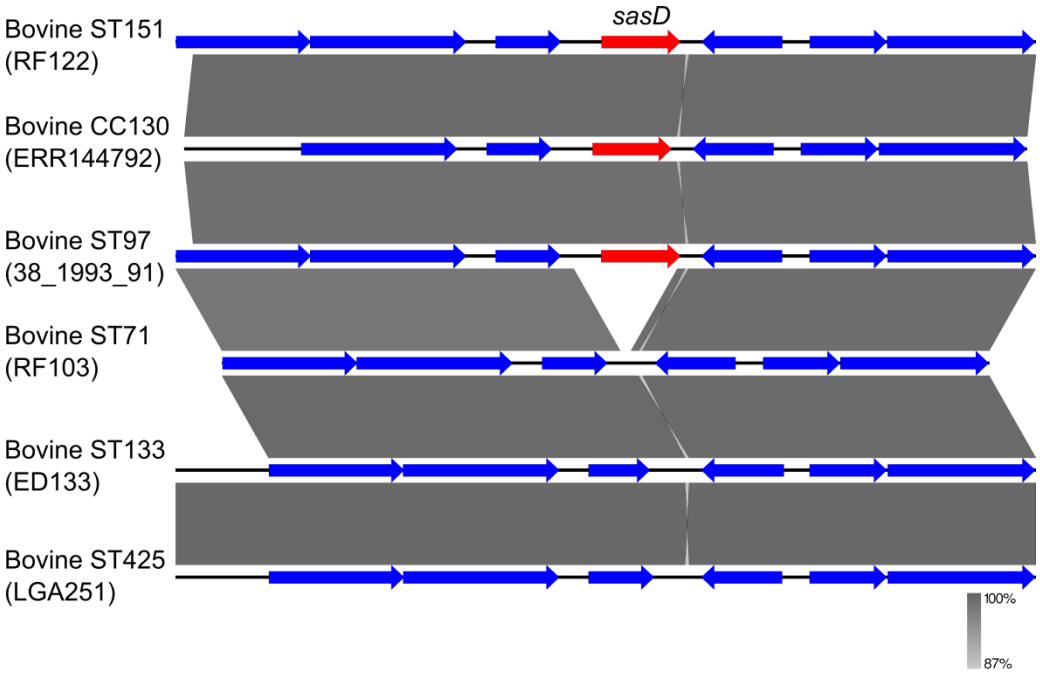


Figure 5.5. Schematic diagram depicting the chromosomal location of the *sasD* gene in different ruminant STs. Representative strains from each ruminant-associated genotype are shown with specific strain names in brackets. The *sasD* gene is labelled red with surrounding genes coloured blue. Pairwise BLASTN sequence comparisons are as shown with the grayscale representing the level of nucleotide identity as shown on the lower right side. Schematic diagram produced using Easyfig (Sullivan, Petty and Beatson, 2011).

A) MW2524(hypothetical protein) **B) MW2526 (malate dehydrogenase)**

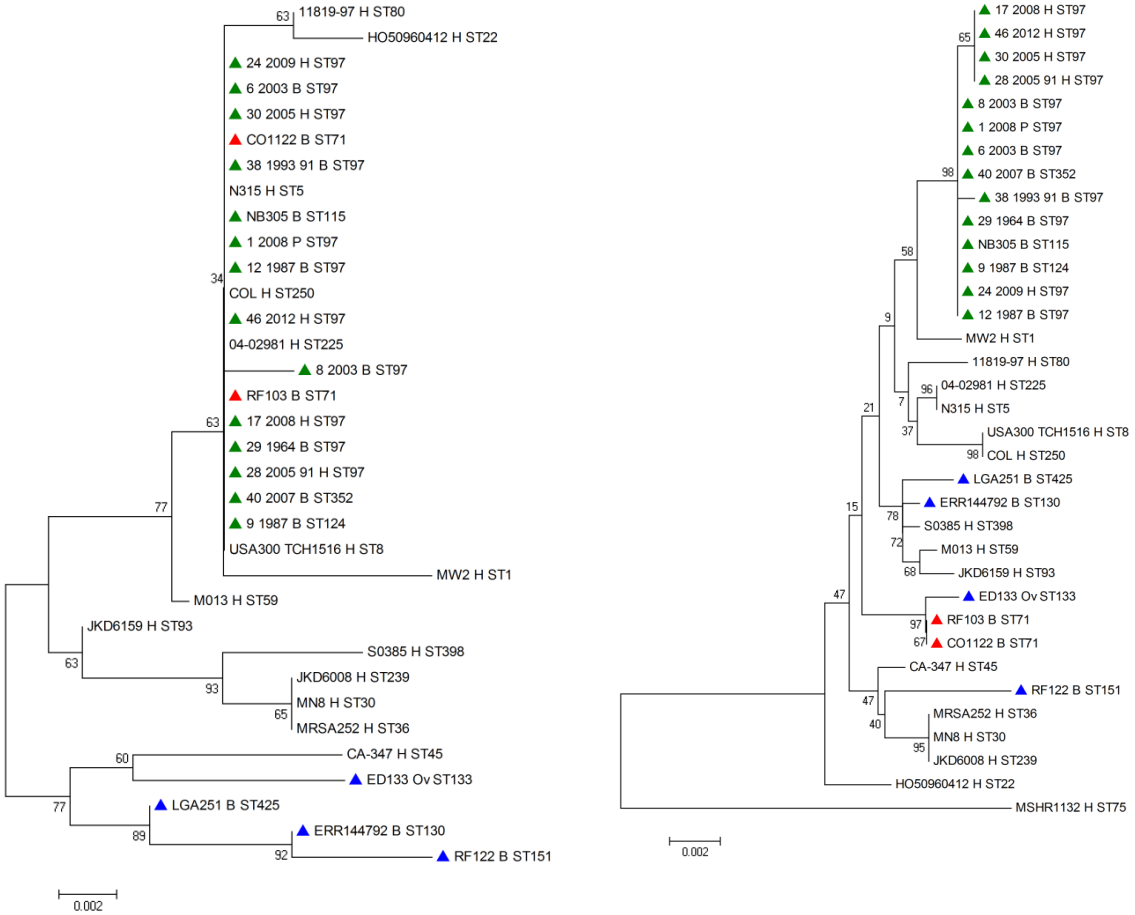


Figure 5.6 (A,B). Neighbour joining trees for A) MW2524 (outwith "SNP dense" region) and B) MW2526 (within "SNP dense" region). Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

C) MW0286 (hypothetical protein) **D) MW0288 (hypothetical protein)**

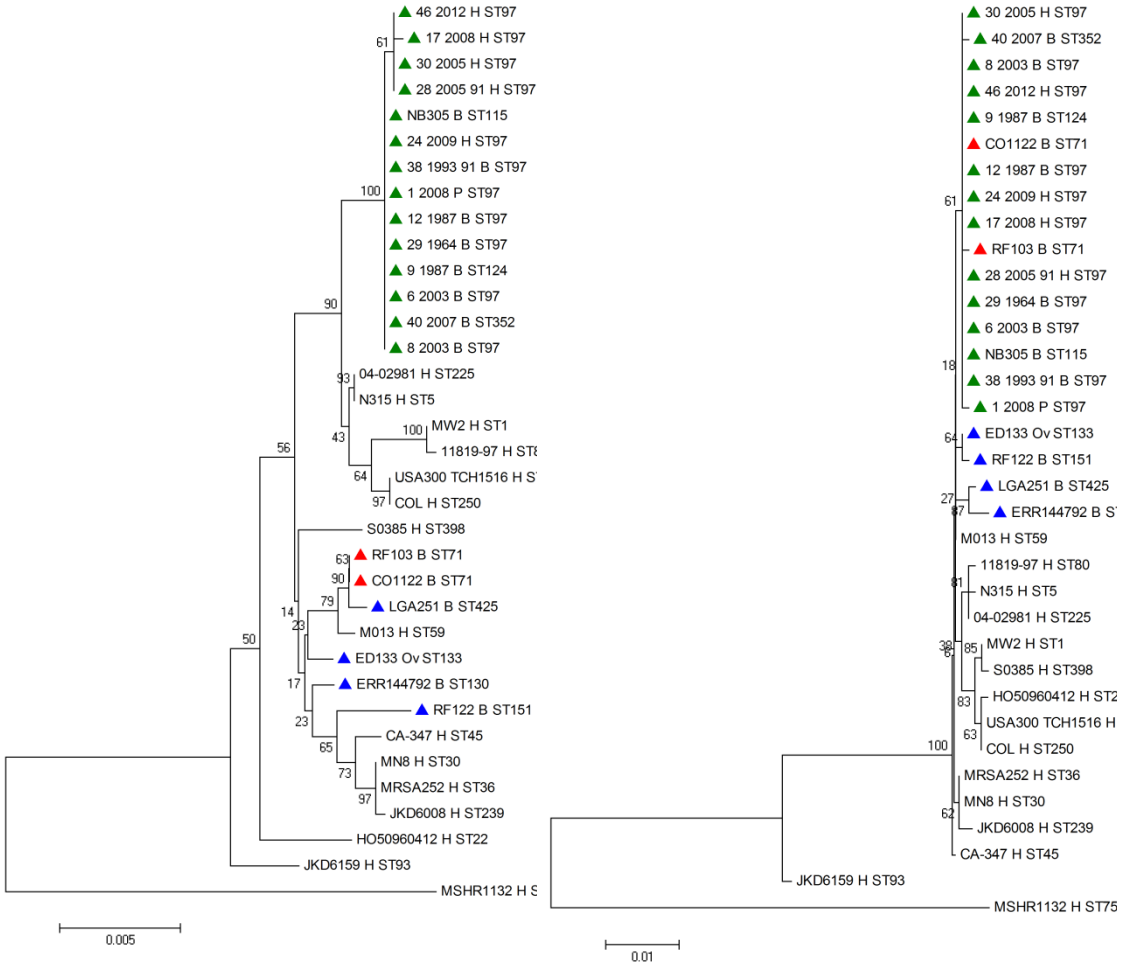


Figure 5.6 (C,D). Neighbour joining trees for C) MW0286 (within "SNP dense" region) and D) MW0288 (outwith "SNP dense" region). Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

5.4.3 Recombination detection analysis reveals multiple predicted recombination events within ST71 *S. aureus*.

To investigate in more detail the nature of the predicted recombination events, including the identification of breakpoints and potential donor sequences, a core genome alignment of the ST71 strains, CC97 strains (including ST97 and other single locus variants) and non-CC97 ruminant *S. aureus* strains was analysed using the recombination detection software BratNextGen (Marttinen et al., 2012). BRATNextGen software has previously been shown to be successful in detecting recombination among bacterial population samples, including *S. aureus* (Castillo-Ramírez et al., 2012).

The software identified the region spanning the origin of replication in the ST71 strains as the result of recombination, indicated by the red segments at either end of the genome alignment (Figure 5.7). The recombinogenic segments identified in strains RF103 and CO1122 are listed in Table 5.5. Among the 2 ST71 strains there were 11 recombinogenic segments identified in both RF103 and CO1122 (Table 5.5). All predicted recombinant regions in both ST71 strains contain overlapping breakpoints and were therefore identified as containing shared origins in the ST71 strains. Of these, 4 segments were predicted to have CC133 strain ED133 as an origin, 1 segment had CC130 strain ERR144792 as an origin, and the remaining segments had an unknown origin, indicating that the donor sequence was not contained in the alignment. These regions totalled 323 080 bp in RF103 and 322800 bp in CO1122 (Table 5.5). There were no recombinant segments identified in the other non-CC97 ruminant strains included in the analysis (ED133, LGA251, RF122, ERR144792), or in any of the CC97 strains (Figure 5.7, Table 5.5).

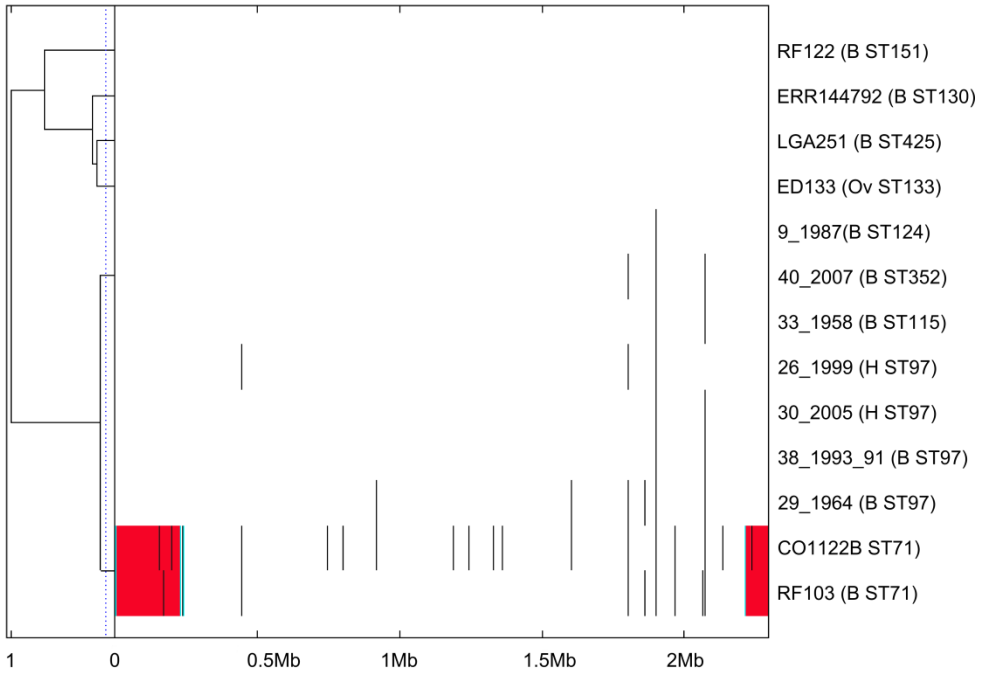


Figure 5.7. Recombination detection in CC97 *S. aureus* and non-CC97 *S. aureus*. The names of each strain are shown on the right hand side with host (B=bovine, Ov=ovine and H=human) and MLST sequence indicated in brackets. On the left side of the diagram is the proportion of shared ancestry tree (PSA) as determined by BratNextGen (Marttinen et al., 2012). On the right side of the diagram the coloured panels indicate detected recombination events along the length of the core genome alignment. A continuous stretch of a single colour indicates a single recombinogenic segment, and the same colour at overlapping genomic locations in different strains indicates those segments are of the same origin. However, the same colour at different genomic locations does not mean the same ancestral origins for those segments. Black lines indicate alignment gaps. For specific details on each segment refer to Table 5.5.

Table 5.5. Recombination detected by BratNextGen in CC97 and non-CC97 *S. aureus* strains

Recombinant strain	Start	End	Size (bp)	Origin
RF103	1	808	807	Unknown
RF103	809	3187	2378	ED133
RF103	3188	227241	224053	Unknown
RF103	227242	229097	1855	ED133
RF103	229098	229839	741	Unknown
RF103	233350	238777	5427	Unknown
RF103	238778	241099	2321	ED133
RF103	2207658	2210411	2753	ED133
RF103	2210412	2252933	42521	Unknown
RF103	2252934	2253612	678	ERR144792
RF103	2253613	2293159	39546	Unknown
CO1122	1	808	807	Unknown
CO1122	809	3187	2378	ED133
CO1122	3188	227141	223953	Unknown
CO1122	227142	229197	2055	ED133
CO1122	229198	229839	641	Unknown
CO1122	233630	238777	5147	Unknown
CO1122	238778	241099	2321	ED133
CO1122	2207658	2210260	2602	ED133
CO1122	2210261	2252933	42672	Unknown
CO1122	2252934	2253612	678	ERR144792
CO1122	2253613	2293159	39546	Unknown

5.4.4. ST71 and ST97 differ in gene content across the predicted recombinant region

In order to characterise further the differences between the ST71 and ST97 strains in the predicted recombinant region, this region was analysed for gene content variation using Cortex (Iqbal et al., 2012).

The capsule HIJK gene cluster is known to specify the capsule type that is produced by *S. aureus* (O'Riordan and Lee, 2004). The ST71 strains had genes which would confer capsule type 8 (Table 5.6), while the ST97 strains had genes encoding capsule type 5 (Table 5.7). In addition, the ST71 strains were both found to contain *cna*, the gene encoding collagen adhesion (CNA), which is a cell wall-anchored (CWA) protein that mediates binding to collagen (Zong et al., 2005) . This gene was absent in all of the bovine ST97 *S. aureus* strains examined.

There were several additional virulence genes absent in the ST71 strains compared to the ST97 strains (Table 5.7). As described earlier (Section 4.4.3) the *sasD* gene was confirmed as absent in both ST71 strains (Table 5.4). Also absent in the ST71 strains were the genes of the *ica* operon (*icaA, icaD, icaB, icaC)* and the transcriptional regulator, *icaR* (Table 5.7). The *ica* operon is responsible for the biosynthesis of the extracellular polysaccharide adhesin known as PIA (polysaccharide intercellular adhesin) involved with biofilm formation (Cramton et al., 1999). In addition, the *his* operon, which is involved in biosynthesis of the amino acid histidine (Alifano et al., 1996), was absent in the ST71 strains but present in the ST97 strains (Table 5.7). Of note, genes encoding a type 1 restriction-modification (RM) system (*hsdS, hsdM, hsdR*) were present in the ST97 strains, but absent in the ST71 strains (Table 5.7). Both ST97 and ST71 strains contain RM systems in genomic islands νSaα and νSaβ, but the additional RM system identified in the ST97 strains was located downstream of *orfX* (Figure 5.11). The RM system is a barrier to horizontal transfer of DNA, in which specific target sequences are recognised and cleaved by an endonuclease,

while a modification protein modifies the same target sequence of "self" DNA to protect it (Waldron and Lindsay, 2006). The target sequences recognised by the RM systems and *hsdS* specificity genes have been shown to be lineage-specific, which may have attributed to the distribution of MGE observed within these strains (Roberts et al., 2013; Waldron and Lindsay, 2006). In place of the RM system, the ST71 strains had a transposase and 5 hypothetical proteins, consistent with a possible novel MGE (Figure 5.11).

Given the difference in the number of genotype-specific CDS observed between the ST71 and ST97 strains, a comparison of the size of the predicted recombinant region was compared between the ST71 strain *de novo* assemblies and orthologous regions in the ST97 strains (Table 5.8, Figure 5.10). Considerable size variation was observed among the strains, although the ST71 strains had the smallest regions, consistent with the results of the gene content analysis (Table 5.6, Table 5.7).

Table 5.6. Protein coding sequences identified to be present in ST71 strains but not ST97 strains

Gene Name	Locus Tag *	Product
cna	SARLGA251_24600	Collagen adhesin precursor
	SARLGA251_02280	Nitric oxide reductase subunit B
cap8H	MW0131	Cap8H capsular polysaccharide synthesis enzyme Cap8H
cap8I	MW0132	Cap8I capsular polysaccharide synthesis enzyme
cap8J	MW0133	Cap8J Capsular polysaccharide synthesis enzyme Cap8J
Cap8K	MW0134	Cap8K capsular polysaccharide synthesis enzyme CapK
	SAB0026	Enterotoxin-like protein
	MW0064	LysR family transcriptional regulator
	SARLGA251_24290	Putative lipoprotein

*Locus tags referred to according to annotations in ST1 human strain MW2(NC_003923) or in the case of core variable genes that are not present in MW2, alternative locus tags from bovine strains RF122 (NC_007622) and LGA251 (FR821779) are listed.

Table 5.7. Protein coding sequences identified to be present in ST97 strains but not ST71 strains

Gene name	Locus_tag**	Product
cap5H	Newbould305_0696	Capsular polysaccharide synthesis protein O-acetyl transferase Cap5H
cap5I	Newbould305_0697	Capsular polysaccharide biosynthesis protein Cap5I
cap5J	Newbould305_0698	Capsular polysaccharide synthesis protein Cap5J
cap5K	Newbould305_0699	Capsular polysaccharide biosynthesis protein
	Newbould305_0716	Formate dehydrogenase
hsdM	Newbould305_0618	Type I restriction-modification system DNA methylase
hsdS	Newbould305_0619	Type I restriction-modification system specificity protein
hsdR	Newbould305_0620	Type I site-specific deoxyribonuclease, HsdR family
hisD	Newbould305_0552	Histidinol dehydrogenase
hisC	Newbould305_0551	Histidinol-phosphate aminotransferase
hisF	Newbould305_0547	Imidazole glycerol phosphate synthase subunit HisF
his1E	Newbould305_0546	Histidine biosynthesis bifunctional protein HisIE
hisH	Newbould305_0549	Imidazole glycerol phosphate synthase subunit HisH
icaA	Newbould305_0539	Intercellular adhesion protein A
icaD	Newbould305_0540	Intercellular adhesion protein D
icaC	Newbould305_0542	Intercellular adhesion protein C
icaB	Newbould305_0541	Polysaccharide intercellular adhesin deacetylase icaB
icaR	Newbould305_0538	Biofilm operon icaABCD HTH-type negative transcriptional regulator IcaR
sasD	Newbould305_0674	Cell wall surface anchor family protein
	Newbould305_0560	DNA-directed RNA polymerase subunit delta
hisA	Newbould305_0548	1-(5-phosphoribosyl)-5-(5-phosphoribosylamino)methylideneaminoimidazole-4-carboxamide isomerase

Gene name	Locus_tag**	Product
hisB	Newbould305_0550	Imidazoleglycerol-phosphate dehydratase
	Newbould305_0616	Guanylate cyclase
hisZ	Newbould305_0554	ATP phosphoribosyltransferase regulatory subunit
	Newbould305_0555	Polysaccharide deacetylase
	Newbould305_0480	Metallo-beta-lactamase
	Newbould305_0633	ATPase
	Newbould305_0558	Cobalt ABC transporter ATP-binding protein
	Newbould305_0544	Lipase
	Newbould305_0633	RNA helicase
	Newbould305_0638	Tandem lipoprotein
	Newbould305_0562	Lactonase drp35
	Newbould305_0533	Methionine sulfoxide reductase A
	Newbould305_0555	Polysaccharide deacetylase
	Newbould305_0632	Membrane spanning protein
	Newbould305_0563	Rhodanese domain sulfurtransferase
	Newbould305_0534	Acetyltransferase
	Newbould305_0564	Pyrrolidone-carboxylate peptidase
	Newbould305_0641	Amidohydrolase
cysG	Newbould305_0486	Precorrin-2 dehydrogenase
	Newbould305_0793	Hexitol dehydrogenase
	Newbould305_0736	RND transporter
	Newbould305_0757	NADH-dependent dehydrogenase
	Newbould305_0815	Ribose transporter RbsU
	Newbould305_0748	Gamma-glutamyltransferase
	Newbould305_0719	4'-phosphopantetheinyl transferase

**Locus tags referred to according to annotations in ST115 (CC97) bovine strain Newbould305 (Accession AKYW01000000).

Table 5.8. Comparison of the size of the predicted recombinant region

Strain	ST	Number of bases in predicted recombinant region
CO1122	71	373910
RF103	71	412475
8_2003	97	436791
40_2007	352	437912
2_2000	97	438274
6_2003	97	438704
5_2003	97	439448
29_1964	97	439547
NB305	115	440189
37_1993_91	97	441611
38_1993_91	97	442428
10_1987	97	444112
15_2006_75	novel SLV	466912
7_2003	97	508508
12_1987	97	511192
33_1958	115	512565
9_1987	124	557650
36_1993_83	97	567788
27_1956	97	592834

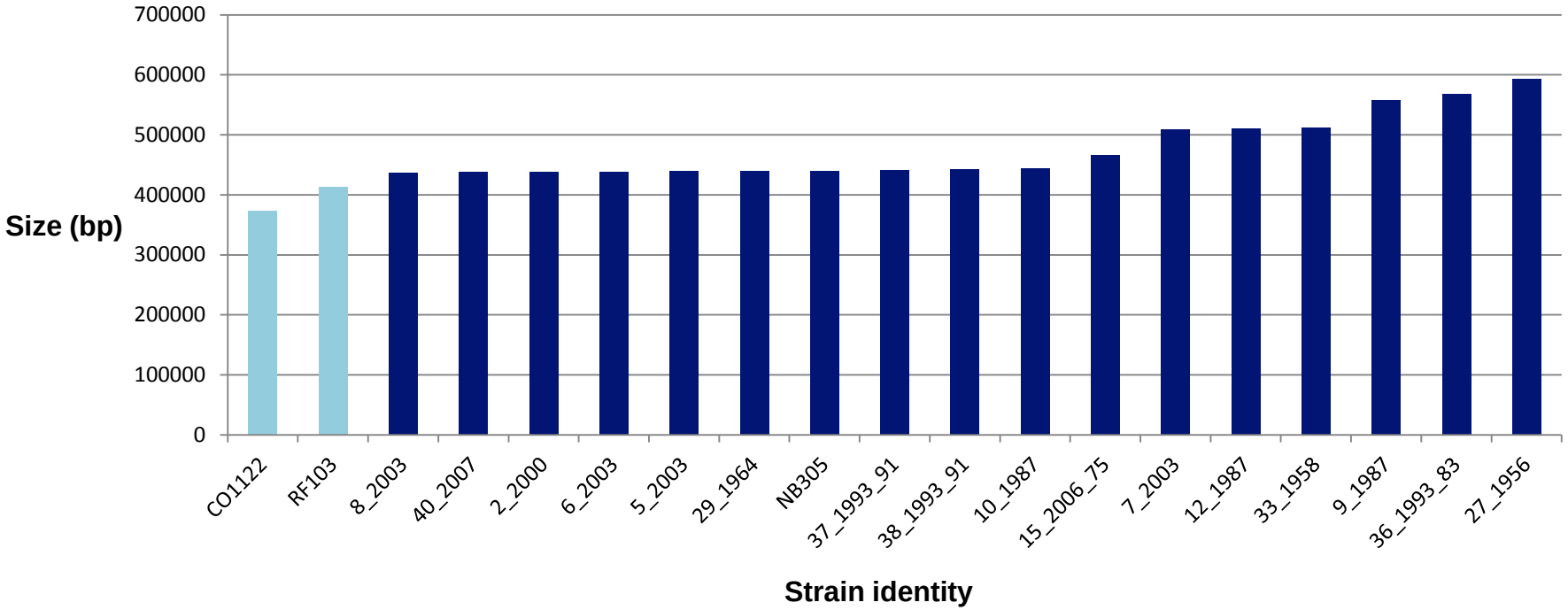


Figure 5.8. Bar graph comparing the size of recombinant region in ST71 and ST97 *S. aureus* strains. Strains are listed along the X axis and size of region in base pairs is indicated on the Y axis.

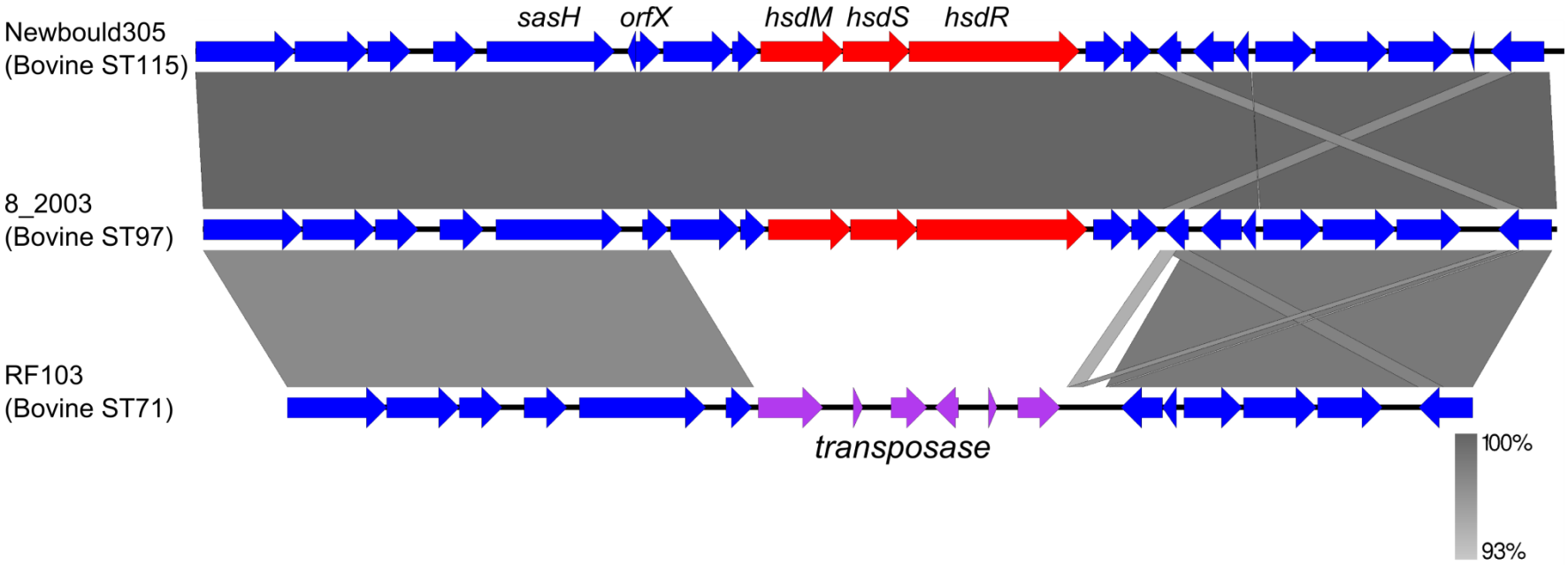


Figure 5.9. Schematic diagram illustrating the region containing the additional type 1 RM system in the ST97 *S. aureus* compared to the ST71 *S. aureus* strains. Representative strains are as labelled on the left hand side of the diagram. Genes of interest are colour-coded (red=type 1 RM system, purple=putative MGE in RF103). Genes are labelled on the diagram where appropriate. The grayscale illustrates the level of nucleotide identity according to BLASTN pairwise sequence comparison.

5.5 Discussion

Homologous recombination events involving large regions of the genome have previously been identified in several bacterial species, including *Escherichia coli* (Touchon et al., 2009), *Streptococcus agalactiae* (Brochet et al., 2008), *Streptococcus pneumoniae* (Hiller et al., 2010) and *Acinetobacter baumanii* (Snitkin et al., 2011). In addition, the evolution of several human-associated lineages of *S. aureus*, including the clinically important HA-MRSA CC239 clone has been attributed to the creation of a mosaic genome content from at least 2 distinct genetic backgrounds as a result of large-scale homologous recombination (Robinson and Enright, 2004). In the current study, the data indicate that large-scale chromosomal replacements have contributed to the diversification of the mastitis-associated ST71 single locus variant from the ancestral genotype, ST97.

The phylogenetic position of the ST71 strains on an extended branch from the remainder of the CC97 *S. aureus* lineage prompted further investigation into the genetic relationship between ST71 and the ST97 *S. aureus* strains. The non-random distribution of densely-clustered SNPs spanning the origin of replication between ST71 and ST97 suggests that this region contains distinct evolutionary origins from the remainder of the genome, and given the distribution, is less likely to represent a hypermutable strain, although these have also been identified in *S. aureus* associated with mastitis (Wang et al., 2013).

The observation that the proportion of synonymous and nonsynonymous SNPs varied greatly within and outwith the "SNP dense" region (Table 5.3) is also consistent with that region having a different evolutionary origin in the ST71 strains. Although dN/dS ratios have not been calculated in the current study, the relative enrichment of synonymous SNPs in the "SNP dense" region of the ST71 strains might indicate homologous recombination has occurred with a strain that is older in evolutionary terms, in which mildly deleterious mutations have been purged over

time. Castillo-Ramirez *et al* observed a relative enrichment of synonymous SNPs in genes in core genome and MGE which had been horizontally transferred in *S. aureus* (Castillo-Ramírez et al., 2011). The elevated frequency of synonymous SNPs observed in the ST71 recombinant region is consistent with the ST97 genotype acquiring gene sequences from a distinct lineage that has been associated with ruminants for a longer time period. Previous phylogenetic analysis indicated that both CC133/CC425 and CC151/CC130 ruminant lineages emerged in ruminants at least 1200 and 5400 years respectively, prior to the emergence of the CC97 lineage (Weinert et al., 2012). Since these genotypes have been associated with the ruminant host for a greater period of time compared to CC97, the horizontal acquisition of an import into the CC97 genetic background from a genotype that theoretically may be more adapted to the ruminant host, may confer a fitness advantage.

Overall, the results of the current study have found evidence of a mosaic genome composition in the bovine ST71 strains spanning the origin of replication, that is of distinct evolutionary origin. In addition, there was not a predominant ruminant-associated genotype identified as the donor sequence, suggesting that there may have been multiple recombination events, or that recombination has occurred in this region among several ruminant genotypes previously, followed by a single horizontal import of a mosaic sequence into an ST97 *S. aureus* strain. In *S. aureus* and other species of Staphylococci such as *S. epidermidis* and *S. haemolyticus*, this region (*oriC* environ) has been observed to have considerable plasticity and may be more susceptible to the horizontal import of exogenous DNA (Noto et al., 2008; Takeuchi et al., 2005). Examination of the border regions in more detail may shed some light onto whether more recombination events have taken place, since there is evidence from the differing right hand side breakpoints observed in ST239 of more than one import involved in the evolution of the ST239 lineage (Robinson and Enright, 2004).

Potential donor genotypes identified by the recombination detection software and phylogenetic analysis of individual genes included non-CC97 ruminant-associated

genotypes such as CC133, CC130, CC151, CC425 or an as yet unknown donor sequence. While the molecular typing and phylogenetic single gene analysis included only a small number of genes, the recombination detection programme involved a core genome-wide analysis, identifying a donor lineage based on a much larger region of sequence, which may explain the variation observed between the 2 methods in the identity of donor sequences in certain regions. Overall, there is not a single defined donor lineage identified suggesting that there has been multiple recombination events which have contributed to the variation. These results contrast with the previously identified hybrid genome of ST239, which was found to have an ST30 donated region and an ST8 background (Robinson and Enright, 2004), and ST34 and ST42, which were found to have an ST10/ST425 donor and an ST30 and ST39 background, respectively (Robinson and Enright, 2004). ST10 and ST425 are single locus variants of each other, differing only at the *yqiL* locus (http://saureus.mlst.net/).

A striking finding is that for each hybrid *S. aureus* genome that has been detected, including ST71, homologous recombination has occurred at a region that always spans the origin of replication (Robinson and Enright, 2004). However the size of the recombinant region varies between lineages. In the current study, the predicted recombinant region of the ST71 strains was at least 24 kbp less than the smallest orthologous region in the bovine ST97 strains (Table 5.8). This was consistent with the smaller number of unique regions identified in the ST71 strains compared to the ST97 strains (Table 5.6, Table 5.7). The size of these regions are intermediate in size when compared to the 557 kbp recombinant region in ST239, and the 244 kbp and 256 kbp regions identified in ST34 and ST42 respectively (Robinson and Enright, 2004). The fact that the ST71 strains appear to have a reduced genome size in this region as a result of the recombination may be suggestive of reductive evolution as an adaptive trait in ruminant *S. aureus*. Reductive evolution has been observed in bacteria undergoing a transition to a niche environment such as an intracellular environment (Moran and Plague, 2004), and it is feasible that the ST71 lineage may be undergoing transition to an intracellular lifestyle within the mammary gland, since

this has been proposed as a cause of chronic subclinical infections (Garzoni and Kelley, 2009). Potential niche environments might include the epithelial cells of the teat skin, or mammary gland alveolar cells (Almeida et al., 1996; Garzoni and Kelley, 2009).

As a result of the large-scale homologous recombination, there were multiple observations of gene loss or gain in this region between the ST71 *S. aureus* and their ancestral lineage ST97. This included genes known to encode virulence proteins or surface-expressed proteins that interact with the host, and thus may give rise to phenotypic differences resulting in different pathogenic potential. Overall there appeared to be an increased number of protein coding sequences that have been lost in the ST71 strains, compared to what has been gained as a result of the recombination. However, using the methods stated and cutoffs for kmer length, which were set in order to increase sensitivity in detecting novel regions of short length (Section 4.3.6), it was not possible to quantify the exact number of hypothetical proteins present in each set of strains, and therefore it is possible, depending on the origin of the import, that there are proportionately more uncharacterised hypothetical proteins in the ST71 strains. Comparison of the size of sequence in the recombinant region of the ST71 and ST97 does suggest that the ST71 overall have a reduced amount of sequence as a result of the recombination, which may be an adaptive trait of reductive evolution (see general discussion, Chapter 6).

The ST71 strains were observed to contain the collagen adhesion gene (*cna*), which was not identified in any of the bovine ST97 strains (Table 5.6). The collagen adhesion gene is a surface-expressed adhesin that has been shown to mediate binding to collagen (Zong et al., 2005), and has been implicated in evasion of host immune defences through inhibition of the classical activation pathway of the complement system (Kang et al., 2013). The *cna* gene has been found to be present in *S. aureus* of bovine origin in approximately 30 % of isolates (Ote et al., 2011; Zecconi et al.,

2006). Further work to investigate the impact of *cna* acquisition on the phenotype of ST71 *S. aureus* strains might involve *in vitro* collagen adherence assays to compare the ability of the strains to bind collagen, and comparison of the virulence of isogenic mutants in experimental infections.

The ST71 and ST97 *S. aureus* strains were shown to contain different capsule specificity genes (*cap*HIJK), consistent with encoding capsule serotype type 8 for ST71 and capsule serotype 5 for ST97. This is consistent with previous findings by Guinane *et al*, using DNA microarray and Southern blot-hybridisation methods to determine capsule type in ST71 and ST97 strains, compared to ST151 (Guinane et al., 2008). Capsular polysaccharides facilitate *S. aureus* in resisting phagocytosis, and are common capsule types described in bovine mastitis-causing strains (Kampen, Tollersrud and Lund, 2005; Poutrel et al., 1988; Tollersrud et al., 2000; Wolf et al., 2011). A switch in capsule phenotype would generate antigenic diversity within an infecting bacterial population and facilitate evolution of the strains in escaping host adaptive immune responses. For example, recombination in genes involved in capsule biosynthesis was a method by which strains of *Streptococcus pneumoniae* have switched capsule serotype on multiple occasions in order to escape selective pressures imposed by vaccines (Croucher et al., 2011).

The *ica* operon (*ica ADBC*) is involved in producing polysaccharide intercellular adhesin (PIA) which is involved in biofilm formation (Cramton et al., 1999), and is highly prevalent among bovine mastitis *S. aureus* isolates (Vasudevan et al., 2003). These genes along with their transcriptional regulator gene *icaR* were found to be absent in the ST71 strains, but were present in all bovine ST97 strains (Table 5.6, 5.7), indicating a potential difference of the ST71 and ST97 strains in their ability to form biofilm. However, *ica*-independent methods of biofilm formation have been identified in *S. aureus*, including cell-to-cell aggregation mediated by the biofilm-associated (Bap) protein (Cucarella et al., 2001), in addition to LPXTG-containing cell wall-associated proteins including the fibronectin-binding proteins (FnBPA and

FnBPB), SasG, SasC and protein A, which also mediate *ica*-independent biofilm formation (Corrigan et al., 2007; Houston et al., 2011; Merino et al., 2009; O'Neill et al., 2008; Schroeder et al., 2009). As such, *S. aureus* strains have been shown to be able to produce biofilm in the absence of a functional *ica* operon (Fitzpatrick, Humphreys and O'Gara, 2005). Therefore, further phenotypic characterisation would be required *in vitro* and *in vivo* to determine whether this genetic difference results in a phenotypic difference in the ability of the ST71 and ST97 *S. aureus* strains to form biofilm.

The absence of the genes involved in histidine biosynthesis in the ST71 *S. aureus* strains suggests that this pathway is inactivated in these strains. This may result in these strains being auxotrophic for this amino acid, instead relying on an exogenous supply of histidine from the surrounding environment. The genome sequence of bovine ST151 strain RF122 (NC_007622) indicates that the *hisC* gene is a pseudogene in this strain, which would also be consistent with a predicted phenotype that is auxotrophic for histidine due to inactivation of the histidine biosynthesis pathway. In a previous study by Delorme *et al*, the histidine requirements of *Lactococcus lactis* from dairy and non-dairy sources was compared (Delorme et al., 1993). It was found that while the majority (91 %) of strains from non-dairy sources were prototrophic for histidine, the majority (93 %) of strains from dairy sources were auxotrophic for histidine, or partly auxotrophic, due to inactivating mutations affecting different parts of the histidine biosynthesis pathway (Delorme et al., 1993). Taken together, these findings indicate that the dairy environment may select for the loss of histidine production (Delorme et al., 1993). It is currently not known whether being auxotrophic for histidine is associated with adaptation to the niche environment of the mammary gland in *S. aureus*. Further work to investigate this would be to conduct *in vitro* experiments using selective media that are deficient in histidine, to compare the ability of ST71 and ST97 strains to grow in the presence and absence of histidine.

The identification of an additional type 1 restriction modification (RM) system in the bovine CC97 strains that was not present in the ST71 strains is consistent with the plasticity of this region, and type 1 RM systems have been unknown to be acquired here (Noto et al., 2008). Although the ST71 strains lack this type 1 RM system, in its place resides a putative MGE of unknown function which may confer alternative virulence properties. Since the type 1 RM system is a barrier to horizontal gene transfer (Waldron and Lindsay, 2006), this genetic difference might account for the increased ability of the ST71 strains to undergo homologous recombination. Further work to investigate this would include comparing the ability of the ST71 and ST97 strains to uptake exogenous DNA such as plasmids, since ST97 has been previously identified as non-transformable with plasmid DNA, although only a single *S. aureus* strain was used in the study (Monk et al., 2012).

Taken together these findings indicate important genetic differences between ST71 and their ancestral genotype, ST97 as a result of homologous recombination events in the region of the origin of replication. The findings from the current study are the first identification of such large-scale recombination events occurring in an animal-adapted lineage of *S. aureus*. This has generated a large amount of genetic diversity, which may generate phenotypic diversity in the CC97 ruminant-associated bacterial population. These results are consistent with large-scale recombination in other bacteria, such as the events identified in the *Acinetobacter baumanni* nosocomial outbreak strains, which also resulted in genetic differences that could affect pathogenic potential, such as in cell-surface antigens and the O-antigen biosynthetic gene cluster (Snitkin et al., 2011).

The current study has characterised large-scale recombination that has shaped the evolution of the ruminant *S. aureus* ST71 from the ST97 ancestor. It is possible that large-scale recombination events involving the core genome are more widespread than previously thought. A recent study by Everitt *et al* identified widespread core genome homologous recombination which has occurred over the long term evolution

of *S. aureus*, with an excess of homoplasies observed among certain clonal complexes including CC239, CC1, CC5, CC8, CC15, CC25 and CC97, although no ST71 strains were included (Everitt et al., 2014). In particular, hotspots of core genome recombination were identified in regions flanking MGE, including the region spanning the origin of replication, suggesting that over the long term, large-scale recombination in the core genome is driven by horizontal transfer of MGE (Everitt et al., 2014). Taken together, the findings of the current study and the study by Everitt *et al*. suggest that recombination in *S. aureus* is more widespread than previously thought and has contributed to the evolution of *S. aureus* over the long term.

6. General Discussion

It has been estimated that 80 % of emerging pathogens of importance to human health are associated with animal reservoirs (Woolhouse and Gaunt, 2007). Of these, the majority appear to be viruses, such as human immunodeficiency virus (HIV-1), the primary aetiological agent of the AIDS pandemic (Keele et al., 2006; Van Heuverswyn et al., 2007). Animal reservoirs have been shown to be an important ecological factor contributing towards the successful transmission of human bacterial pathogens. For example, rodent populations facilitated transmission of the bacterial agent of the plague, *Yersinia pestis* (Cui et al., 2013). Bacteria including *S. aureus* have been known to undergo host jumps from humans to animals (Eppinger et al., 2006; Guinane et al., 2010; Lowder et al., 2009; Weinert et al., 2012). The increased resolution and practicality afforded by high throughput sequencing methods is enabling researchers to address questions relating to population-based datasets of important microbial pathogens. To date, many of these studies have focussed upon clones of importance to human health only (Harris et al., 2010; 2013; McAdam et al., 2012). With the increasing problem in treating multi-resistant microbial infections, an understanding of microbial evolution, particularly in relation to the ability to occupy diverse ecological niches remains important in light of designing effective interventions such as vaccination or biosecurity control measures.

In the current study, the evolutionary history of *S. aureus* clone CC97 was investigated. At least 2 independent livestock-to-human host jumps have occurred in recent times, identifying livestock as an origin for a clone of human-adapted CC97 CA-MRSA that has spread globally. Livestock have also been identified as the origin for the emergence of human *S. aureus* clones CC59 and CC25 approximately 500 years ago (Weinert et al., 2012). However, although MLST is useful for studying the evolution of *S. aureus* over long time periods, the limited resolution of the sequence data is not useful for identifying more recent host jumps, such as the events found in the current study. The findings of the current study indicate that a lineage that has long been associated with a ruminant host is still capable of jumping into humans, and that host jump events may occur on more than one occasion within a lineage, prompting concerns for multiple emergence events of human-associated subclones.

The data from the current study indicates that the original ancestor of the CC97 lineage was human-associated, which is consistent with studies conducted on other ruminant-associated lineages (Guinane et al., 2010; Herron-Olson et al., 2007). The Bayesian credibility intervals for the estimated time of emergence for the CC97 lineage from a predicted human ancestor in this study overlap with those previously estimated using MLST sequence data, with both estimates indicating that CC97 has been associated with cattle for approximately 1000 years (Weinert et al., 2012). This clone is therefore predicted to have emerged in cattle more recently than the other major bovine clones CC133 and CC151 which were estimated to have emerged approximately 3000 years and 5000 years ago, respectively (Weinert et al., 2012). In contrast, the CC97 livestock-to-human host jumps may have occurred as recently as 40 years ago, which highlights the need to carry out further whole genome sequencing studies of other livestock-associated clones, in order to identify other host switches that may have occurred in the recent past. For example, there are other *S. aureus* clones that colonise livestock and also infect other host species for which the evolutionary origins have not yet been established. For example, CC130 is one of several clones containing the novel *mecC* methicillin resistance gene, which has a wide distribution in humans in several countries in Europe including the UK, Ireland, Germany, Denmark and France (Cuny et al., 2011; García-Álvarez et al., 2011; Shore, et al., 2011; Stegger et al., 2012). Epidemiological evidence based on CC130 spa type t843 indicated an increasing prevalence in human MRSA infections between 2003-2011 (Petersen et al., 2013). Zoonotic transmission of CC130 between cows and humans has been identified (Harrison et al., 2013), and it is feasible that livestock-to-human host jumps may also occur, highlighting the need for future research into the evolutionary origins of this clone, and other clones with multiple host tropisms.

The transmission dynamics and epidemiological factors involved in the successful emergence of a pathogen in a novel host are likely to be multi-factorial, relating to the molecular features of the pathogen, the host and the surrounding environment.

The intensification of agricultural systems including high stocking densities, frequent use of antimicrobials, selective breeding and nutrition that pushes for the highest milk yields at the lowest cost may all provide selective pressures that drive the evolution and diversification of bacteria into a new host. It is likely that repeated or prolonged contact between humans and livestock have occurred in order for a successful host switch of *S. aureus* CC97 to occur. This might inform future biosecurity measures and herd health management with the aim of decreasing contact between livestock and humans in the dairy setting, and minimising the spread of mastitis within the dairy herd, both of which may minimise subsequent transmission events between livestock and humans. Examples of on-farm biosecurity measures might include maintaining a closed herd or isolation of new arrivals to the farm, adequate maintenance of milking liners and other milking equipment to minimise trauma to the teat skin at milking and colonisation of milk liners by bacteria, and the use of gloves and strict disinfection practices during milking to minimise contact between human skin and contaminated farm or livestock surfaces.

Currently the majority of information related to *S. aureus* on the DEFRA website pertains to the control of human nosocomial infections (https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs). The results of the current study suggest that the public health focus should also be directed towards livestock, and as such awareness to be increased in the veterinary and agricultural industry. This might include implementing increased surveillance procedures in the agricultural setting, as has been done in human medicine in the UK, as surveillance of MRSA and MSSA bacteremia cases is mandatory for all NHS acute trusts (http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/StaphylococcusAureus/EpidemiologicalData/MandatorySurveillance/). European legislation states that random testing of bulk tank milk samples is carried out regularly for antibiotic residues (Food Standards Agency, 2009). Theoretically, additional mandatory microbiological testing of such milk samples might increase early detection of potentially pathogenic *S. aureus* clones for humans, such as CC97.

Another potential route by which humans might come into contact with livestock-associated *S. aureus* clones might be through the contamination of food products of animal origin. *S. aureus* is known to be a common contaminant of food products such as meat from beef, lamb, pork and poultry (De Boer et al., 2009; Fessler, et al., 2011) in addition to products made from milk such as cheese and ice cream (Jakobsen et al., 2011; Reinoso et al., 2008). Since millions of people buy and consume meat products, it is feasible that contamination of kitchen surfaces with *S. aureus* during food preparation might provide another route by which humans may be exposed to CC97 *S. aureus* that is capable of switching hosts into humans. *S. aureus* is known to survive on surfaces in the home for several months, and CC97 *S. aureus* has been isolated from the surfaces of houses, the source of which was not identified (Roberts et al., 2011b). The prevalence of CC97 *S. aureus* on food products of animal origin is not currently known. A single CC97 *S. aureus* isolate has so far been described in a cheese product made from raw milk in Norway (Jørgensen et al., 2005). Further work would be required to determine the prevalence of CC97 *S. aureus* contamination of animal products and thereby assess the risk for human exposure by this route. Often bacteriological food studies focus on detection of MRSA (De Boer et al., 2009; Fessler et al., 2011) and the results of the current study indicate that the prevalence of MSSA should also be examined, as all bovine CC97 isolates were sensitive to methicillin.

Following the livestock-to-human host jump, CC97 human clade A has disseminated globally, with closely-related CC97 *S. aureus* distributed in several countries on 4 different continents, which is most likely to have been as a result of the globalised nature of air travel. Intercontinental air travel has been implicated in the dissemination of many pathogens including the coronavirus that caused the SARS epidemic, Influenza and antimicrobial-resistant Enterobacteriaceae (Brownstein, Wolfe and Mandl, 2006; Olsen et al., 2003; Van der Bij and Pitout, 2012; Wilder-Smith, Leong and Villacian, 2006). Additional sampling and genome sequencing of CC97 human *S. aureus* strains over time will be required to determine whether there

is any subsequent phylogeographic clustering of the isolates in each country, such as that observed with other clones such as CC239 (Harris et al., 2010).

In contrast human clade B appears to be more representative of either a single outbreak in the Midlands region of the UK or a clade in the very early stages of clonal expansion. The UK-based human clade B CC97 strains uniquely have the ACME element. ACME is known to contribute to infection through promoting growth and survival in conditions that mimic those of human skin, in addition to promoting persistence by inhibiting factors affecting wound healing (Thurlow et al., 2013). The presence of ACME is widely associated with the epidemic USA300 lineage, and has contributed towards the success of this clone by increasing survival and transmission (Diep et al., 2008). The presence of this clone in the human-associated CC97 background may therefore provide a fitness advantage analogous to USA300. This warrants further sampling to determine whether this clone has the potential for further clonal expansion and global spread.

The current study also identified molecular features of MGE that were associated with host species from which CC97 strains were isolated. In particular, following the livestock-to-human host jumps, the human CC97 subclones had acquired genes on MGE which conferred antimicrobial resistance and are involved in the evasion of human innate immune defenses. The human CC97 subclones also lacked MGE typically associated with ruminants. These results highlight the central role that horizontal transfer of MGE plays in the adaptation of strains to a particular host. While other studies document MGE changes associated with a jump from humans into livestock (Guinane et al., 2010; Herron-Olson et al., 2007; Lowder et al., 2009; Price et al., 2012), this is the first study to document those adaptive traits after recent host jump events from an animal back into the human host. Identification of molecular features associated with a transition between different host species may provide some potential targets for intervention strategies such as vaccination. If one can understand how a clone can adapt to a new host following a host jump, then

targeting the mechanisms used to do this may be effective in preventing the successful adaptation and emergence of future clones in a new host species. For example, examination of the distribution of the MGE in human and poultry ST5 found that the poultry strains contained MGE that were specific to the avian host (Lowder et al., 2009).

The findings of the current study also indicate that the CC97 strains have MGE features that correlate with host specificity. In theoretical considerations for vaccine design, targeting host-specific determinants seems a logical idea for preventing *S. aureus* from becoming established in a new ecological niche should a host jump occur. However, due to the fact that MGE may be acquired and lost frequently, a vaccine designed against an antigen contained on a MGE may not be successful. The only vaccine currently targeted at an antigen contained on a MGE is a subunit vaccine against PVL which is phage-encoded (Karauzum et al., 2013). So far, a successful vaccine that prevents *S. aureus* infection in humans has not yet been found. In humans, the primary aim is to prevent infection, and vaccine targets have largely comprised surface expressed general *S. aureus* antigens such as capsule, wall teichoic acid, ClfA, and IsdB among others (Fowler and Proctor, 2014). The reason for the vaccines failing, often at human clinical trial level is not known and may be complex. However, it has recently been proposed by some authors that since *S. aureus* is a commensal that has evolved with humans over thousands of years, and it is known to produce multiple proteins involved in evading immune defenses, that these proteins should be targeted for vaccines, including phage-encoded SAK SCN and CHIPS as well as the anti-opsonin SpA (Fowler and Proctor, 2014; Kobayashi and DeLeo, 2013). The findings of the current study suggest that acquisition of MGE encoding immune evasion proteins appears to be important in adaptation of CC97 *S. aureus* to the human host. However, the complement of IEC genes in human *S. aureus* varies and is not contained in every single strain (Wamel et al., 2006), indicating that these might not be the optimal antigens to use as a vaccine target.

Since livestock are a reservoir for *S. aureus* known to be important for human public health, an alternative vaccine strategy may be to design a ruminant vaccine that

prevents or decreases ruminant infection to a level at which transmission to humans becomes much less likely. So far, bovine vaccines have included various combinations of bacterins (inactivated bacterial strains), toxoid and adjuvants, or DNA-based and recombinant protein vaccines, all with varying degrees of efficacy (Pereira et al., 2011). The findings of this study indicate that continuing research on bovine vaccine design is warranted. Not only would this benefit the dairy industry greatly in reducing the burden and huge costs associated with dealing with mastitis, but it may also decrease the livestock reservoir of *S. aureus*, thus potentially preventing further zoonotic transmissions and host jump events.

The current study identified a striking difference in the antimicrobial susceptibility profiles of the bovine isolates compared to the pig and human isolates. While the pig and human CC97 *S. aureus* were often found to be resistant to multiple classes of antimicrobial, all of the bovine CC97 were MSSA, with only resistance to penicillin identified. This implies that the dairy industry has not played a major role in the development of resistance and that in the case of CC97, human strains acquired resistance after transitioning to the human host. The reason behind these differences is not currently known and may reflect factors relating to the pathogen, host or differing antimicrobial selective pressures encountered between the pig and dairy industry. For example, one explanation might relate to a separation between the bacteria and antimicrobial treatment in the bovine host in some way, which may also explain why cure rates are highly variable (Barkema et al., 2009). This might be caused by pathogen factors such as the ability of the strain to form biofilm or persist intracellularly within the mammary gland (Cucarella et al., 2004; Garzoni and Kelley, 2009; Vasudevan et al., 2003). Host factors might involve the differing immune systems and selective breeding programs between pigs and cattle. For example, in chronic mammary infections fibrosis and leukocyte infiltration might alter the structure of the mammary gland, forming abscesses that effectively wall off the bacteria, preventing penetration by antimicrobials (Akers and Nickerson, 2011). Environmental factors might relate to differing veterinary prescription of antimicrobials between the pig and dairy industry. For example, in pigs, while growth-promoting sub-therapeutic doses of antimicrobials in feed have been banned

in the EU since 2006, antimicrobials such as chlortetracycline in feed premixes are still extensively prescribed for prophylactic and therapeutic indications such as enzoonotic pneumonia caused by respiratory pathogens such *as Mycoplasma hyopneumoniae*, or *Pasteurella multocida.* The evolution of CC398 LA-MRSA is thought to have involved the acquisition of tetracycline and methicillin resistance by strains which evolved from a human MSSA ancestor (Price et al., 2012). In contrast, in the dairy industry, notwithstanding dry cow therapy and systemic injections, treatment of mastitis in lactating cows is typically by administration of an intramammary preparation of antimicrobial such as β-lactams for several days. These antimicrobials are not predicted to penetrate intracellularly and cure rates can be variable (Sol et al., 2000). The recent emergence of *mecC*-positive MRSA in bovine-associated genotypes such as CC130 and CC425 still implies that bacteria are continuing to evolve and diversify to acquire resistance against antimicrobials in the dairy setting (García-Álvarez et al., 2011; Shore, et al., 2011). In the current study, none of the CC97 *S. aureus* strains contained this element. However this warrants further research to determine the exact origin of the *mecC* gene and whether it will continue to spread to other ruminant clones such as CC97, CC133 and CC151.

In the current study, a large-scale recombination event involving a region of the core genome spanning the origin of replication was found in the ST71 mastitis-associated clone. While mosaic regions of distinct evolutionary origins involving such a large region of the genome has previously been described in several human-associated clones such as CC239 (Robinson and Enright, 2004), the observations for ST71 are the first description of such large-scale recombination occurring in a clone associated with ruminants, suggesting that recombination may play a role in adaptive diversification in the ruminant host also. Large-scale recombination events have also been described in other bacteria and are proposed to be an adaptive mechanism that rapidly generates large-scale diversity in a bacterial population (Brochet et al., 2008; Hiller et al., 2010; Snitkin et al., 2011). In *S. aureus* which is considered to be highly clonal in population structure, recombination of DNA with already-resident flora in a particular ecological niche may represent a method of rapidly generating genetic

diversity that may facilitate adaptation by altering the phenotype and fitness of a bacterial population. In all large-scale recombination events identified in *S. aureus* to date, the region involved varies in size but in all examples spans the origin of replication. The region just downstream of the origin of replication has been termed the *oriC* environ, which demonstrates high plasticity when compared among different Staphylococcal species such as *S. aureus, S. haemolyticus* and *S. epidermidis* (Takeuchi et al., 2005). Even within *S. aureus*, this region has been noted as prone to the acquisition of exogenous genetic elements (Noto et al., 2008). The fact that this has now been found in several strains from animal and human hosts raises the question as to how widespread large-scale recombination is across the general *S. aureus* population, in particular in different host species such as humans, livestock, companion animals, rabbits and birds. Recent research indicates that core genome recombination has contributed to the long-term evolution of several human-associated clones of *S. aureus*, which may be driven by the horizontal transfer of MGE, that simultaneously transfers flanking regions of core genome (Everitt et al., 2014). This identifies an area for further research in order to determine the prevalence and role of large-scale recombination in the evolution of wider *S. aureus* natural populations from a variety of host species.

In addition to the areas of further research already stated, a useful extension to this study would be to assess how the genotypic differences observed relate to changes in phenotype that would potentially affect the fitness of the strains in different hosts. For example, the fitness and virulence of CC97 *S. aureus* isolated from different hosts might be examined under a variety of growth conditions that might mimic the different environments encountered in each host, using a phenotype microarray approach such as Biolog (http://www.biolog.com/). Many of the MGE associated with host specificity in CC97 have as yet uncharacterised genes that may play a role in host adaptation. The function of such genes may be assessed by creating isogenic knockout strains for a particular MGE followed by subsequent complementation of individual genes and using appropriate *in vitro* assays to determine the role of individual genes contained on each MGE. In the case of the ST71 *S. aureus* strains,

the role that the large scale recombination has had in affecting the phenotype of the strains compared to the ancestral ST97 genotype might be compared using *in vitro* assays that compare certain differences in gene content. For example, the differential ability of ST71 and ST97 isolates in binding collagen might be assessed based on the presence of the *cna* gene in the ST71 strains.

To summarise, this study has identified the evolutionary origins of a livestock-associated clone, identifying livestock-to-human host jumps that have resulted in a clone of CA-MRSA that has disseminated worldwide. This therefore highlights the animal-livestock interface as a continued important area for research in order to identify as early as possible any future host jumps and emerging clones, and for the design of appropriate control measures. In addition, the molecular features of host adaptation associated with the recent host jumps into humans were identified. Finally, the identification of large-scale recombination in an animal-associated lineage opens up questions about how important recombination is in the adaptive evolution of *S. aureus*, and provides a basis for future research efforts.

References

Aarestrup, F. M., Cavaco, L., and Hasman, H. (2010). Decreased susceptibility to zinc chloride is associated with methicillin resistant *Staphylococcus aureus* CC398 in Danish swine. *Vet. Microbiol.*, *142*(**3-4**), 455–7.

Abdelbary, M. M. H., Wittenberg, A., Cuny, C., Layer, F., Kurt, K., Wieler, L. H., Walther, B., Skov, R., Larsen, J., Hasman, H., Fitzgerald, J. R., Smith, T. C., Wagenaar, J. A., Pantosti, A., Hallin, M., Struelens, M. J., Edwards, G., Böse, R., Nübel, U., and Witte, W. (2014). Phylogenetic analysis of *Staphylococcus aureus* CC398 reveals a sub-lineage epidemiologically associated with infections in horses. *PLoS One*, *9*(**2**), e88083.

Aires-de-Sousa, M., Conceição, T., and de Lencastre, H. (2006). Unusually high prevalence of nosocomial Panton-Valentine leukocidin-positive *Staphylococcus aureus* isolates in Cape Verde Islands. *J. Clin. Microbiol.*, *44*(**10**), 3790–3.

Aires-de-Sousa, M., Parente, C. E. S. R., Vieira-da-Motta, O., Bonna, I. C. F., Silva, D. A., and de Lencastre, H. (2007). Characterization of *Staphylococcus aureus* isolates from buffalo, bovine, ovine, and caprine milk samples collected in Rio de Janeiro State, Brazil. *Appl. Environ. Microbiol.*, *73*(**12**), 3845–3849.

Akers, R. M., and Nickerson, S. C. (2011). Mastitis and its impact on structure and function in the ruminant mammary gland. *J. Mammary Gland Biol. Neoplasia*, *16*(**4**), 275–89.

Albrecht, N., Jatzwauk, L., Slickers, P., Ehricht, R., and Monecke, S. (2011). Clonal Replacement of Epidemic Methicillin-Resistant *Staphylococcus aureus* Strains in a German University Hospital over a Period of Eleven Years. *PLoS One*, *6*(**11**), e28189.

Alder, M., and Easton, G. (2005). Human and veterinary medicine. *BMJ*, *330*(**7496**), 858–9.

Alifano, P., Fani, R., Lio, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M.S., and Bruni, C.B. (1996). Histidine biosynthetic pathway and genes: structure, regulation and evolution. *Microbiol. Rev.*,*60*(**1**), 44-69.

Alikhan, N.-F., Petty, N. K., Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, **(12)**, 402.

Almeida, R. A., Matthews, K. R., Cifrian, E., Guidry, A. J., and Oliver, S. P. (1996). *Staphylococcus aureus* invasion of bovine mammary epithelial cells. *J. Dairy Sci.*, *79*(**6**), 1021–6.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, *215*(**3**), 403–410.

Anderson, M. E. C., Lefebvre, S. L., Rankin, S. C., Aceto, H., Morley, P. S., Caron, J. P., Welsh, R. D., Holbrook, T. C., Moore, B., Taylor, D. R., and Weese, J. S. (2009). Retrospective multicentre study of methicillin-resistant *Staphylococcus aureus* infections in 115 horses. *Equine Vet. J.*, *41*(**4**), 401–405.

Armand-Lefevre, L., Ruimy, R., and Andremont, A. (2005). Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. *Emerg. Infect. Dis.*, *11*(**5**), 711–714.

Atalla, H., Gyles, C., Wilkie, B., Leslie, K., and Mallard, B. (2009). Somatic cell scores and clinical signs following experimental intramammary infection of dairy cows with a *Staphylococcus aureus* small colony variant (*S. aureus* SCV) in comparison to other bovine strains. *Vet. Microbiol.*, *137*(**3–4**), 326–334.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*, 75.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., and Hiramatsu, K. (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet*, *359*(**9320**), 1819–1827.

Bager, F., Madsen, M., Christensen, J., and Aarestrup, F. . (1997). Avoparcin used as a growth promoter is associated with the occurrence of vancomycin-resistant *Enterococcus faecium* on Danish poultry and pig farms. *Prev. Vet. Med.*, *31*(**1-2**), 95–112.

Baptiste, K. E., Williams, K., Willams, N. J., Wattret, A., Clegg, P. D., Dawson, S., Corkill, J. E., O'Neill, T., and Hart, C. A. (2005). Methicillin-resistant Staphylococci in companion animals. *Emerg. Infect. Dis.*, *11*(**12**), 1942–1944.

Barkema, H. W., Green, M. J., Bradley, A. J., and Zadoks, R. N. (2009). Invited review: The role of contagious disease in udder health. *J. Dairy Sci.*, *92*(**10**), 4717–4729.

Barkema, H. W., Schukken, Y. H., Lam, T. J., Beiboer, M. L., Wilmink, H., Benedictus, G., and Brand, A. (1998). Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. *J. Dairy Sci.*, *81*(**2**), 411–9.

Barrio, M. B., Rainard, P., and Prévost, G. (2006). LukM/LukF'-PV is the most active *Staphylococcus aureus* leukotoxin on bovine neutrophils. *Microbes Infect.*, *8*(**8**), 2068–74.

Bartels, M. D., Boye, K., Rhod Larsen, A., Skov, R., and Westh, H. (2007). Rapid increase of genetically diverse methicillin-resistant *Staphylococcus aureus*, Copenhagen, Denmark. *Emerg. Infect. Dis.*, *13*(**10**), 1533–1540.

Basic-Hammer, N., Vogel, V., Basset, P., and Blanc, D. S. (2010). Impact of recombination on genetic variability within *Staphylococcus aureus* clonal complexes. *Infect. Genet. Evol.*, *10*(**7**), 1117–1123.

Bates, J., Jordens, J. Z., and Griffiths, D. T. (1994). Farm animals as a putative reservoir for vancomycin-resistant enterococcal infection in man. *J. Antimicrob. Chemother.*, *34*(**4**), 507–514.

Battisti, A., Franco, A., Merialdi, G., Hasman, H., Iurescia, M., Lorenzetti, R., Feltrin, F., Zini, M., and Aarestrup, F. M. (2010). Heterogeneity among methicillin-resistant *Staphylococcus aureus* from Italian pig finishing holdings. *Vet. Microbiol.*, *142*(**3–4**), 361–366.

Begier, E. M., Frenette, K., Barrett, N. L., Mshar, P., Petit, S., Boxrud, D. J., Watkins-Colwell, K., Wheeler, S., Cebelinski, E. A., Glennen, A., Nguyen, D., Hadler, J. L., and Team, T. C. B. F. E. R. (2004). A High-Morbidity Outbreak of Methicillin-Resistant *Staphylococcus aureus* among Players on a College Football Team, Facilitated by Cosmetic Body Shaving and Turf Burns. *Clin. Infect. Dis.*, *39*(**10**), 1446–1453.

Ben Zakour, N. L., Sturdevant, D. E., Even, S., Guinane, C. M., Barbey, C., Alves, P. D., Cochet, M.-F., Gautier, M., Otto, M., Fitzgerald, J. R., and Le Loir, Y. (2008). Genome-wide analysis of ruminant *Staphylococcus aureus* reveals diversification of the core genome. *J. Bacteriol.*, *190*(**19**), 6302–6317.

Berg, T., Firth, N., Apisiridej, S., Hettiaratchi, A., Leelaporn, A., and Skurray, R. A. (1998). Complete Nucleotide Sequence of pSK41: Evolution of Staphylococcal Conjugative Multiresistance Plasmids. *J. Bacteriol.*, *180*(**17**), 4350–4359.

Berglund, C., Ito, T., Ikeda, M., Ma, X. X., Söderquist, B., and Hiramatsu, K. (2008). Novel type of staphylococcal cassette chromosome *mec* in a methicillin-resistant *Staphylococcus aureus* strain isolated in Sweden. *Antimicrob. Agents Chemother.*, *52*(**10**), 3512–6.

Bhatty, M., Laverde Gomez, J. A., and Christie, P. J. (2013). The expanding bacterial type IV secretion lexicon. *Res. Microbiol.*, *164*(**6**), 620–39.

Boakes, E., Kearns, A. M., Ganner, M., Perry, C., Warner, M., Hill, R. L., and Ellington, M. J. (2011). Molecular diversity within clonal complex 22 methicillin-resistant *Staphylococcus aureus* encoding Panton-Valentine leukocidin in England and Wales. *Clin. Microbiol. Infect.*, *17*(**2**), 140–5.

Bouchard, D., Peton, V., Almeida, S., Maréchal, C. Le, Miyoshi, A., Azevedo, V., Berkova, N., Rault, L., François, P., Schrenzel, J., Even, S., Hernandez, D., and

Loir, Y. Le. (2012). Genome Sequence of *Staphylococcus aureus* Newbould 305, a Strain Associated with Mild Bovine Mastitis. *J. Bacteriol.*, *194*(**22**), 6292–6293.

Boundy, S., Safo, M. K., Wang, L., Musayev, F. N., O'Farrell, H. C., Rife, J. P., and Archer, G. L. (2013). Characterization of the *Staphylococcus aureus* rRNA methyltransferase encoded by *orfX*, the gene containing the staphylococcal chromosome Cassette *mec* (SCC*mec*) insertion site. *J. Biol. Chem.*, *288*(**1**), 132–40.

Brochet, M., Rusniok, C., Couvé, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., and Glaser, P. (2008). Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc. Natl. Acad. Sci.*, *105*(**41**), 15961 –15966.

Brosch, R., Gordon, S. V, Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D., and Cole, S. T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U. S. A.*, *99*(**6**), 3684–9.

Brownstein, J. S., Wolfe, C. J., and Mandl, K. D. (2006). Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Med.*, *3*(**10**), e401.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.*, *68*(**3**), 560–602.

Burke, F., McCormack, N., Rindi, S., Speziale, P., and Foster, T. (2010). Fibronectin-binding protein B variation in *Staphylococcus aureus*. *BMC Microbiol.*, *10*(**1**), 160.

Burland, T. G. (2000). DNASTAR's Lasergene sequence analysis software. *Methods Mol. Biol.*, *132*, 71-91.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brüssow, H. (2003). Prophage genomics. *Microbiol. Mol. Biol. Rev.*, *67*(**2**), 238–76.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., and Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics*, *21*(**16**), 3422–3.

Castillo-Ramírez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W. P., Westh, H., Boye, K., Gulay, Z., Bentley, S. D., Parkhill, J., Holden, M. T., and Feil, E. J. (2012). Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol.*, *13*(**12**), R126.

Castillo-Ramírez, S., Harris, S. R., Holden, M. T. G., He, M., Parkhill, J., Bentley, S. D., and Feil, E. J. (2011). The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog.*, *7*(**7**), e1002129.

Cavaco, L. M., Hasman, H., and Aarestrup, F. M. (2011). Zinc resistance of *Staphylococcus aureus* of animal origin is strongly associated with methicillin resistance. *Vet. Microbiol.*, *150*(**3-4**), 344–348.

Cavaco, L. M., Hasman, H., Stegger, M., Andersen, P. S., Skov, R., Fluit, A. C., Ito, T., and Aarestrup, F. M. (2010). Cloning and occurrence of *czrC*, a gene conferring cadmium and zinc resistance in methicillin-resistant *Staphylococcus aureus* CC398 isolates. *Antimicrob. Agents Chemother.*, *54*(**9**), 3605–8.

Chan, K. S., Ling, M. L., Hsu, L. Y., and Tan, A. L. (2009). Methicillin-resistant *Staphylococcus aureus* throat colonization among healthcare workers during an outbreak in Singapore General Hospital. *Infect. Control Hosp. Epidemiol. Off. J. Soc. Hosp. Epidemiol. Am.*, *30*(**1**), 95–97.

Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., Shah, S., Rudrik, J. T., Pupp, G. R., Brown, W. J., and others. (2003). Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. *N. Engl. J. Med.*, *348*(**14**), 1342.

Chatzakis, E., Scoulica, E., Papageorgiou, N., Maraki, S., Samonis, G., and Galanakis, E. (2011). Infant colonization by *Staphylococcus aureus*: role of maternal carriage. *Eur. J. Clin. Microbiol. Infect. Dis.*, *30*(**9**), 1111–1117.

Chongtrakool, P., Ito, T., Ma, X. X., Kondo, Y., Trakulsomboon, S., Tiensasitorn, C., Jamklang, M., Chavalit, T., Song, J.-H., and Hiramatsu, K. (2006). Staphylococcal cassette chromosome *mec* (SCC*mec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCC*mec* elements. *Antimicrob. Agents Chemother.*, *50*(**3**), 1001–12.

Chua, K., Seemann, T., Harrison, P. F., Davies, J. K., Coutts, S. J., Chen, H., Haring, V., Moore, R., Howden, B. P., and Stinear, T. P. (2010). Complete Genome Sequence of *Staphylococcus aureus* Strain JKD6159, a Unique Australian Clone of ST93-IV Community Methicillin-Resistant *Staphylococcus aureus*, *192*(**20**), 5556–5557.

Chung, M., Dickinson, G., De Lencastre, H., and Tomasz, A. (2004). International clones of methicillin-resistant *Staphylococcus aureus* in two hospitals in Miami, Florida. *J. Clin. Microbiol.*, *42*(**2**), 542–547.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).*, *6*(**2**), 80–92.

Clarke, S. R., Wiltshire, M. D., and Foster, S. J. (2004). IsdA of *Staphylococcus aureus* is a broad spectrum, iron-regulated adhesin. *Mol. Microbiol.*, *51*(**5**), 1509–1519.

Clinical and Laboratory Standards Institute. (2007). Performance Standards for Antimicrobial Susceptibility Testing. CLSI document M100-S17.

Coello, R., Glynn, J. R., Gaspar, C., Picazo, J. J., and Fereres, J. (1997). Risk factors for developing clinical infection with methicillin-resistant *Staphylococcus aureus* (MRSA) amongst hospital patients initially only colonized with MRSA. *J. Hosp. Infect.*, *37*(**1**), 39–46.

Coker, R., Rushton, J., Mounier-Jack, S., Karimuribo, E., Lutumba, P., Kambarage, D., Pfeiffer, D. U., Stärk, K., and Rweyemamu, M. (2011). Towards a conceptual framework to support one-health research for policy on emerging zoonoses. *Lancet Infect. Dis.*, *11*(**4**), 326–31.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(**6823**), 1007–11.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S. R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., and Gagneux, S. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.*, *45*, 1176–1182.

Conceição, T., Sousa, M. A. de, and Lencastre, H. de. (2009). Staphylococcal Interspersed Repeat Unit Typing of *Staphylococcus aureus*: Evaluation of a New Multilocus Variable-Number Tandem-Repeat Analysis Typing Method. *J. Clin. Microbiol.*, *47*(**5**), 1300–1308.

Concepción Porrero, M., Hasman, H., Vela, A. I., Fernández-Garayzábal, J. F., Domínguez, L., and Aarestrup, F. M. (2012). Clonal diversity of *Staphylococcus aureus* originating from the small ruminants goats and sheep. *Vet. Microbiol.*, *156*(**1-2**), 157–161.

Corrigan, R. M., Rigby, D., Handley, P., and Foster, T. J. (2007). The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. *Microbiology*, *153*(**Pt 8**), 2435–46.

Corvaglia, A. R., François, P., Hernandez, D., Perron, K., Linder, P., and Schrenzel, J. (2010). A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proc. Natl. Acad. Sci.*, *107*(**26**), 11954–11958.

Couto, N., Tilley, P., Simões, J., Sales Luis, J. P., and Pomba, C. (2012). First Report of Methicillin-Resistant *Staphylococcus aureus* ST5 and ST398 from Purebred Lusitano Horses. *J. Equine Vet. Sci.*, *32*(**5**), 300–304.

Cramton, S. E., Gerke, C., Schnell, N. F., Nichols, W. W., and Götz, F. (1999). The Intercellular adhesion (*ica*) locus is present in *Staphylococcus aureus* and is required for biofilm formation. *Infect. Immun.*, *67*(**10**), 5427–5433.

Croucher, N. J., Harris, S. R., Fraser, C., Quail, M. A., Burton, J., van der Linden, M., McGee, L., von Gottberg, A., Song, J. H., Ko, K. S., Pichon, B., Baker, S., Parry, C. M., Lambertsen, L. M., Shahinas, D., Pillai, D. R., Mitchell, T. J., Dougan, G., Tomasz, A., Klugman, K. P., Parkhill, J., Hanage, W. P., and Bentley, S. D. (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science*, *331*(**6016**), 430–4.

Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I., and Penadés, J. R. (2001). Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *J. Bacteriol.*, *183*(**9**), 2888–96.

Cucarella, C., Tormo, M. A., Ubeda, C., Trotonda, M. P., Monzon, M., Peris, C., Amorena, B., Lasa, I., and Penades, J. R. (2004). Role of Biofilm-Associated Protein Bap in the Pathogenesis of Bovine *Staphylococcus aureus*. *Infect. Immun.*, *72*(**4**), 2177–2185.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L. A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Li, Y., Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, J., and Yang, R. (2013). Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(**2**), 577–82.

Cuny, C., Friedrich, A., Kozytska, S., Layer, F., Nübel, U., Ohlsen, K., Strommenger, B., Walther, B., Wieler, L., and Witte, W. (2010). Emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) in different animal species. *Int. J. Med. Microbiol.*, *300*(**2-3**), 109–117.

Cuny, C., Kuemmerle, J., Stanek, C., Willey, B., Strommenger, B., and Witte, W. (2006). Emergence of MRSA infections in horses in a veterinary hospital: strain characterisation and comparison with MRSA from humans. *Euro Surveill.*, *11*(**1**), 44–7.

Cuny, C., Layer, F., Strommenger, B., and Witte, W. (2011). Rare occurrence of methicillin-resistant *Staphylococcus aureus* CC130 with a novel *mecA* homologue in humans in Germany. *PLoS One*, *6*(**9**), e24360.

Darling, A. E., Mau, B., and Perna, N. T. (2010). ProgressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One*, *5*(**6**), e11147.

Daum, R. S., Ito, T., Hiramatsu, K., Hussain, F., Mongkolrattanothai, K., Jamklang, M., and Boyle-Vavra, S. (2002). A Novel Methicillin-Resistance Cassette in Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Isolates of Diverse Genetic Backgrounds. *J. Infect. Dis.*, *186*(**9**), 1344 –1347.

David, M. Z., Cadilla, A., Boyle-Vavra, S., and Daum, R. S. (2014). Replacement of HA-MRSA by CA-MRSA Infections at an Academic Medical Center in the Midwestern United States, 2004-5 to 2008. *PLoS One*, *9*(**4**), e92760.

Day, N. P., Moore, C. E., Enright, M. C., Berendt, A. R., Smith, J. M., Murphy, M. F., Peacock, S. J., Spratt, B. G., and Feil, E. J. (2001). A link between virulence and ecological abundance in natural populations of *Staphylococcus aureus*. *Science*, *292*(**5514**), 114–116.

De Boer, E., Zwartkruis-Nahuis, J. T. M., Wit, B., Huijsdens, X. W., de Neeling, A. J., Bosch, T., van Oosterom, R. A. A., Vila, A., and Heuvelink, A. E. (2009). Prevalence of methicillin-resistant *Staphylococcus aureus* in meat. *Int. J. Food Microbiol.*, *134*(**1-2**), 52–6.

De Haas, C. J. C., Veldkamp, K. E., Peschel, A., Weerkamp, F., Van Wamel, W. J. B., Heezius, E. C. J. M., Poppelier, M. J. J. G., Van Kessel, K. P. M., and van Strijp, J. A. G. (2004). Chemotaxis inhibitory protein of *Staphylococcus aureus*, a bacterial antiinflammatory agent. *J. Exp. Med.*, *199*(**5**), 687–95.

De Neeling, A. J., van den Broek, M. J. M., Spalburg, E. C., van Santen-Verheuvel, M. G., Dam-Deisz, W. D. C., Boshuizen, H. C., van de Giessen, A. W., van Duijkeren, E., and Huijsdens, X. W. (2007). High prevalence of methicillin resistant *Staphylococcus aureus* in pigs. *Vet. Microbiol.*, *122*(**3-4**), 366–72.

DeLeo, F. R., Kennedy, A. D., Chen, L., Wardenburg, J. B., Kobayashi, S. D., Mathema, B., Braughton, K. R., Whitney, A. R., Villaruz, A. E., Martens, C. A., Porcella, S. F., McGavin, M. J., Otto, M., Musser, J. M., and Kreiswirth, B. N. (2011). Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proc. Natl. Acad. Sci.*, *108*(**44**), 18091-18096.

Delorme, C., Godon, J.J., Ehrlich, S.D., and Renault, P. (1993). Gene inactivation in *Lactococcus lactis*: histidine biosynthesis. *J. Bacteriol*., 175(**14**), 4391-4399.

Devriese, L. A. (1984). A simplified system for biotyping *Staphylococcus aureus* strains isolated from animal species. *J. Appl. Bacteriol.*, *56*(**2**), 215–220.

Devriese, L. A., Van Damme, L. R., and Fameree, L. (1972). Methicillin (cloxacillin)-resistant *Staphylococcus aureus* strains isolated from bovine mastitis cases. *Zentralbl. Veterinarmed. B*, *19*(**7**), 598–605.

Diekema, D. J., Pfaller, M. A., Schmitz, F. J., Smayevsky, J., Bell, J., Jones, R. N., and Beach, M. (2001). Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillanc. *Clin. Infect. Dis.*, *32*, 114–132.

Diep, B. A., Gill, S. R., Chang, R. F., Phan, T. H. ., Chen, J. H., Davidson, M. G., Lin, F., Lin, J., Carleton, H. A., Mongodin, E. F., and others. (2006). Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet*, *367*(**9512**), 731–739.

Diep, B. A., Perdreau-Remington, F., and Sensabaugh, G. F. (2003). Clonal characterization of *Staphylococcus aureus* by multilocus restriction fragment typing, a rapid screening approach for molecular epidemiology. *J. Clin. Microbiol.*, *41*(**10**), 4559.

Diep, B. A., Stone, G. G., Basuino, L., Graber, C. J., Miller, A., des Etages, S.-A., Jones, A., Palazzolo-Ballance, A. M., Perdreau-Remington, F., Sensabaugh, G. F., DeLeo, F. R., and Chambers, H. F. (2008). The arginine catabolic mobile element and staphylococcal chromosomal cassette *mec* linkage: convergence of virulence and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. *J. Infect. Dis.*, *197*(**11**), 1523–30.

Donker, G. A., Deurenberg, R. H., Driessen, C., Sebastian, S., Nys, S., and Stobberingh, E. E. (2009). The population structure of *Staphylococcus aureus* among general practice patients from The Netherlands. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.*, *15*(**2**), 137–143.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol*, *4*(**5**), e88.

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.*, *7*(**1**), 214.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.*, *29*(**8**), 1969–1973.

Dukic, V. M., Lauderdale, D. S., Wilder, J., Daum, R. S., and David, M. Z. (2013). Epidemics of community-associated methicillin-resistant *Staphylococcus aureus* in the United States: a meta-analysis. *PLoS One*, *8*(**1**), e52722.

Dziekan, G., Hahn, A., Thüne, K., Schwarzer, G., Schäfer, K., Daschner, F. D., and Grundmann, H. (2000). Methicillin-resistant *Staphylococcus aureus* in a teaching hospital: investigation of nosocomial transmission using a matched case-control study. *J. Hosp. Infect.*, *46*(**4**), 263–70.

EARS-Net. (2012). Antimicrobial resistance surveillance in Europe 2012 *(European Antimicrobial Resistance Surveillance Network)* (pp. 59–62).

EFSA. (2010). Analysis of the baseline survey on the prevalence of methicillin-resistant *Staphylococcus aureus* (MRSA) in holdings with breeding pigs, in the EU, 2008 - Part B: factors associated with MRSA contamination of holdings. *EFSA J. 2010*, *8*(**6**), 1–67.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., and Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(**5910**), 133–8.

Ellington, M. J., Hope, R., Livermore, D. M., Kearns, A. M., Henderson, K., Cookson, B. D., Pearson, A., and Johnson, A. P. (2010). Decline of EMRSA-16 amongst methicillin-resistant *Staphylococcus aureus* causing bacteraemias in the UK between 2001 and 2007. *J. Antimicrob. Chemother.*, *65*(**3**), 446–8.

Ellington, M. J., Yearwood, L., Ganner, M., East, C., and Kearns, A. M. (2008). Distribution of the ACME-*arcA* gene among methicillin-resistant *Staphylococcus aureus* from England and Wales. *J. Antimicrob. Chemother.*, *61*(**1**), 73.

Enright, M. C., Day, N. P. ., Davies, C. E., Peacock, S. J., and Spratt, B. G. (2000). Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.*, *38*(**3**), 1008.

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., and Spratt, B. G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc. Natl. Acad. Sci. U. S. A.*, *99*(**11**), 7687–7692.

Eppinger, M., Baar, C., Linz, B., Raddatz, G., Lanz, C., Keller, H., Morelli, G., Gressmann, H., Achtman, M., and Schuster, S. C. (2006). Who Ate Whom? Adaptive *Helicobacter* Genomic Changes That Accompanied a Host Jump from Early Humans to Large Felines. *PLoS Genet*, *2*(**7**), e120.

Ericsson Unnerstad, H., Lindberg, A., Persson Waller, K., Ekman, T., Artursson, K., Nilsson-Ost, M., and Bengtsson, B. (2009). Microbial aetiology of acute clinical mastitis and agent-specific risk factors. *Vet. Microbiol.*, *137*(**1-2**), 90–7.

European Food Safety Authority (EFSA). (2009). Analysis of the baseline survey on the prevalence of methicillin-resistant *Staphylococcus aureus* (MRSA) in holdings with breeding pigs, in the EU, 2008. *EFSA J.*, *7*(**11**), 1376.

European Medicines Agency. (2009). Revised reflection paper on the use of 3rd and 4th generation cephalosporins in food producing animals in the European Union: Development of resistance and impact on human and animal health., 1–37.

Everitt, R. G., Didelot, X., Batty, E. M., Miller, R. R., Knox, K., Young, B. C., Bowden, R., Auton, A., Votintseva, A., Larner-Svensson, H., Charlesworth, J., Golubchik, T., Ip, C. L. C., Godwin, H., Fung, R., Peto, T. E. A., Walker, A. S., Crook, D. W., and Wilson, D. J. (2014). Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. *Nat. Commun.*, **(5)** 3956.

Fang, H., Hedin, G., Li, G., and Nord, C. E. (2008). Genetic diversity of community-associated methicillin-resistant *Staphylococcus aureus* in southern Stockholm, 2000–2005. *Clin. Microbiol. Infect.*, *14*(**4**), 370–376.

Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., Moore, C. E., and Day, N. P. J. (2003). How clonal is *Staphylococcus aureus*? *J. Bacteriol.*, *185*(**11**), 3307–3316.

Feil, E. J., Enright, M. C., and Spratt, B. G. (2000). Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Res. Microbiol.*, *151*(**6**), 465–469.

Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., Zhou, J., and Spratt, B. G. (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. U. S. A.*, *98*(**1**), 182–187.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., and Spratt, B. G. (2004). EBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.*, *186*(**5**), 1518–1530.

Feil, E. J., Maiden, M. C., Achtman, M., and Spratt, B. G. (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol. Biol. Evol.*, *16*(**11**), 1496–1502.

Fessler, A., Scott, C., Kadlec, K., Ehricht, R., Monecke, S., and Schwarz, S. (2010). Characterization of methicillin-resistant *Staphylococcus aureus* ST398 from cases of bovine mastitis. *J. Antimicrob. Chemother.*, *65*(**4**), 619–625.

Fessler, A. T., Kadlec, K., Hassel, M., Hauschild, T., Eidam, C., Ehricht, R., Monecke, S., and Schwarz, S. (2011). Characterization of Methicillin-resistant *Staphylococcus aureus* Isolates from Food and Food Products of Poultry Origin in Germany. *Appl. Environ. Microbiol.*, AEM.00561–11.

Fessler, A. T., Kadlec, K., and Schwarz, S. (2011). Novel Apramycin Resistance Gene apmA in Bovine and Porcine Methicillin-Resistant *Staphylococcus aureus* ST398 Isolates. *Antimicrob. Agents Chemother.*, *55*(**1**), 373–375.

Fitzgerald, J. R. (2012). Livestock-associated *Staphylococcus aureus*: origin, evolution and public health threat. *Trends Microbiol.*, *20*(**4**), 192–8.

Fitzgerald, J. R., Meaney, W. J., Hartigan, P. J., Smyth, C. J., and Kapur, V. (1997). Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. *Epidemiol. Infect.*, *119*(**02**), 261–269.

Fitzgerald, J. R., Monday, S. R., Foster, T. J., Bohach, G. A., Hartigan, P. J., Meaney, W. J., and Smyth, C. J. (2001). Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *J. Bacteriol.*, *183*(**1**), 63–70.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R., and Musser, J. M. (2001). Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc. Natl. Acad. Sci. U. S. A.*, *98*(**15**), 8821–8826.

Fitzpatrick, F., Humphreys, H., and O'Gara, J. P. (2005). Evidence for *icaADBC*-independent biofilm development mechanism in methicillin-resistant *Staphylococcus aureus* clinical isolates. *J. Clin. Microbiol.*, *43*(**4**), 1973–6.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, *269*(**5223**), 496–512.

Food and Drug Administration. (2010). Summary report on antimicrobials used, sold or distributed for use in food-producing animals, 1–5.

Food Standards Agency. (2009). Information and guidance on the testing of milk for antibiotic residues (pp. 1–20).

Fossum Moen, A. E., Šaltytė Benth, J., Alm-Kristiansen, K., and Bukholm, G. (2009). Exotoxin-encoding gene content in community-acquired and hospital-acquired methicillin-resistant *Staphylococcus aureus*. *Clin. Microbiol. Infect.*, *15*(**12**), 1139–1145.

Foster, T. J. (2005). Immune evasion by staphylococci. *Nat. Rev. Microbiol.*, *3*(**12**), 948–958.

Fowler, V. G., and Proctor, R. A. (2014). Where does a *Staphylococcus aureus* vaccine stand? *Clin. Microbiol. Infect.*, *20 Suppl 5*, 66–75.

Francis, J. S., Doherty, M. C., Lopatin, U., Johnston, C. P., Sinha, G., Ross, T., Cai, M., Hansel, N. N., Perl, T., Ticehurst, J. R., Carroll, K., Thomas, D. L., Nuermberger, E., and Bartlett, J. G. (2005). Severe community-onset pneumonia in healthy adults caused by methicillin-resistant *Staphylococcus aureus* carrying the Panton-Valentine leukocidin genes. *Clin. Infect. Dis. An Off. Publ. Infect. Dis. Soc. Am.*, *40*(**1**), 100–107.

Franco, A., Hasman, H., Iurescia, M., Lorenzetti, R., Stegger, M., Pantosti, A., Feltrin, F., Ianzano, A., Porrero, M. C., Liapi, M., and Battisti, A. (2011). Molecular characterization of *spa* type t127, sequence type 1 methicillin-resistant *Staphylococcus aureus* from pigs. *J. Antimicrob. Chemother.*, *66*(**6**), 1231–5.

Francois, P., Huyghe, A., Charbonnier, Y., Bento, M., Herzig, S., Topolski, I., Fleury, B., Lew, D., Vaudaux, P., Harbarth, S., van Leeuwen, W., van Belkum, A., Blanc, D. S., Pittet, D., and Schrenzel, J. (2005). Use of an Automated Multiple-Locus, Variable-Number Tandem Repeat-Based Method for Rapid and High-Throughput Genotyping of *Staphylococcus aureus* Isolates. *J. Clin. Microbiol.*, *43*(**7**), 3346–3355.

Frazee, B. W., Lynn, J., Charlebois, E. D., Lambert, L., Lowery, D., and Perdreau-Remington, F. (2005). High prevalence of methicillin-resistant *Staphylococcus aureus* in emergency department skin and soft tissue infections. *Ann. Emerg. Med.*, *45*(**3**), 311–320.

Fridkin, S. K., Hageman, J., McDougal, L. K., Mohammed, J., Jarvis, W. R., Perl, T. M., and Tenover, F. C. (2003). Epidemiological and microbiological characterization of infections caused by *Staphylococcus aureus* with reduced susceptibility to vancomycin, United States, 1997-2001. *Clin. Infect. Dis.*, *36*(**4**), 429–39.

Furuse, Y., Suzuki, A., and Oshitani, H. (2010). Origin of measles virus: divergence from rinderpest virus between the 11th and 12th centuries. *Virol. J.*, *7*(**1**), 52.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., and Conesa, A. (2012). Qualimap:

evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*(**20**), 2678–9.

García-Álvarez, L., Holden, M. T., Lindsay, H., Webb, C. R., Brown, D. F., Curran, M. D., Walpole, E., Brooks, K., Pickard, D. J., Teale, C., Parkhill, J., Bentley, S. D., Edwards, G. F., Girvan, E. K., Kearns, A. M., Pichon, B., Hill, R. L., Larsen, A. R., Skov, R. L., Peacock, S. J., Maskell, D. J., and Holmes, M. A. (2011). Meticillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect. Dis.*, *11*(**8**), 595–603.

García-Garrote, F., Cercenado, E., Marín, M., Bal, M., Trincado, P., Corredoira, J., Ballesteros, C., Pita, J., Alonso, P., and Vindel, A. (2014). Methicillin-resistant *Staphylococcus aureus* carrying the *mecC* gene: emergence in Spain and report of a fatal case of bacteraemia. *J. Antimicrob. Chemother.*, *69*(**1**), 45–50.

Garcia-Graells, C., Antoine, J., Larsen, J., Catry, B., Skov, R., and Denis, O. (2012). Livestock veterinarians at high risk of acquiring methicillin-resistant *Staphylococcus aureus* ST398. *Epidemiol. Infect.*, *140*(**3**), 383–9.

Garcia-Graells, C., van Cleef, B. A. G. L., Larsen, J., Denis, O., Skov, R., and Voss, A. (2013). Dynamic of livestock-associated methicillin-resistant *Staphylococcus aureus* CC398 in pig farm households: a pilot study. *PLoS One*, *8*(**5**), e65512.

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V, and Hewinson, R. G. (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U. S. A.*, *100*(**13**), 7877–82.

Garzoni, C., and Kelley, W. L. (2009). *Staphylococcus aureus*: new evidence for intracellular persistence. *Trends Microbiol.*, *17*(**2**), 59–65.

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A. A., Zhu, G., Epstein, J. H., Mazet, J. K., Hu, B., Zhang, W., Peng, C., Zhang, Y.-J., Luo, C.-M., Tan, B., Wang, N., Zhu, Y., Crameri, G., Zhang, S.-Y., Wang, L.-F., Daszak, P., and Shi, Z.-L. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, *503*(**7477**), 535–8.

Gharsa, H., Ben Slama, K., Lozano, C., Gómez-Sanz, E., Klibi, N., Ben Sallem, R., Gómez, P., Zarazaga, M., Boudabous, A., and Torres, C. (2012). Prevalence, antibiotic resistance, virulence traits and genetic lineages of *Staphylococcus aureus* in healthy sheep in Tunisia. *Vet. Microbiol.*, (**0**).

Ghasemzadeh-Moghaddam, H., Ghaznavi-Rad, E., Sekawi, Z., Yun-Khoon, L., Aziz, M. N., Hamat, R. A., Melles, D. C., van Belkum, A., Shamsudin, M. N., and Neela, V. (2011). Methicillin-susceptible *Staphylococcus aureus* from clinical

and community sources are genetically diverse. *Int. J. Med. Microbiol. IJMM*, *301*(**4**), 347–353.

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J., Khouri, H., Utterback, T., Lee, C., Dimitrov, G., Jiang, L., Qin, H., Weidman, J., Tran, K., Kang, K., Hance, I. R., Nelson, K. E., and Fraser, C. M. (2005). Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing Methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.*, *187*(**7**), 2426–2438.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, *11*(**8**), R86.

Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., Bröker, B. M., Doskar, J., and Wolz, C. (2009). Diversity of Prophages in Dominant *Staphylococcus aureus* Clonal Lineages. *J. Bacteriol.*, *191*(**11**), 3462 –3468.

Goerke, C., Wirtz, C., Flückiger, U., and Wolz, C. (2006). Extensive phage dynamics in *Staphylococcus aureus* contributes to adaptation to the human host during infection. *Mol. Microbiol.*, *61*(**6**), 1673–85.

Goerke, C., and Wolz, C. (2004). Regulatory and genomic plasticity of *Staphylococcus aureus* during persistent colonization and infection. *Int. J. Med. Microbiol.*, *294*(**2-3**), 195–202.

Golding, G. R., Bryden, L., Levett, P. N., McDonald, R. R., Wong, A., Graham, M. R., Tyler, S., Van Domselaar, G., Mabon, P., Kent, H., Butaye, P., Smith, T. C., Kadlec, K., Schwarz, S., Weese, S. J., and Mulvey, M. R. (2012). Whole-genome sequence of livestock-associated ST398 methicillin-resistant *Staphylococcus aureus* Isolated from Humans in Canada. *J. Bacteriol.*, *194*(**23**), 6627–8.

Golding, G. R., Levett, P. N., McDonald, R. R., Irvine, J., Quinn, B., Nsungu, M., Woods, S., Khan, M., Ofner-Agostini, M., and Mulvey, M. R. (2011). High rates of *Staphylococcus aureus* USA400 infection, Northern Canada. *Emerg. Infect. Dis.*, *17*(**4**), 722–725.

Gómez, P., González-Barrio, D., Benito, D., García, J. T., Viñuela, J., Zarazaga, M., Ruiz-Fons, F., and Torres, C. (2014). Detection of methicillin-resistant *Staphylococcus aureus* (MRSA) carrying the *mecC* gene in wild small mammals in Spain. *J. Antimicrob. Chemother.*, dku100.

Gonzalez, B. E., Hulten, K. G., Dishop, M. K., Lamberth, L. B., Hammerman, W. A., Mason, E. O., and Kaplan, S. L. (2005). Pulmonary manifestations in

children with invasive community-acquired *Staphylococcus aureus* infection. *Clin. Infect. Dis.*, *41*(**5**), 583–90.

Graffunder, E. M. (2002). Risk factors associated with nosocomial methicillin-resistant *Staphylococcus aureus* (MRSA) infection including previous use of antimicrobials. *J. Antimicrob. Chemother.*, *49*(**6**), 999–1005.

Graveland, H., Wagenaar, J. A., Bergs, K., Heesterbeek, H., and Heederik, D. (2011). Persistence of Livestock Associated MRSA CC398 in Humans Is Dependent on Intensity of Animal Contact. *PLoS One*, *6*(**2**), e16830.

Graveland, H., Wagenaar, J. A., Heesterbeek, H., Mevius, D., van Duijkeren, E., and Heederik, D. (2010). Methicillin Resistant *Staphylococcus aureus* ST398 in Veal Calf Farming: Human MRSA Carriage Related with Animal Antimicrobial Usage and Farm Hygiene. *PLoS One*, *5*(**6**), e10990.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S. M., and Poon, L. L. M. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*, *302*(**5643**), 276–8.

Guinane, C. M., Zakour, N. L., Tormo-Mas, M. A., Weinert, L. A., Lowder, B. V., Cartwright, R. A., Smyth, D. S., Smyth, C. J., Lindsay, J. A., Gould, K. A., Witney, A., Hinds, J., Bollback, J. P., Rambaut, A., Penades, J. R., and Fitzgerald, J. R. (2010). Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome Biol. Evol.*, *2*, 454–466.

Guinane, C. M., Sturdevant, D. E., Herron-Olson, L., Otto, M., Smyth, D. S., Villaruz, A. E., Kapur, V., Hartigan, P. J., Smyth, C. J., and Fitzgerald, J. R. (2008). Pathogenomic analysis of the common bovine *Staphylococcus aureus* clone (ET3): emergence of a virulent subtype with potential risk to public health. *J. Infect. Dis.*, *197*(**2**), 205–213.

Haenni, M., Saras, E., Châtre, P., Médaille, C., Bes, M., Madec, J.-Y., and Laurent, F. (2012). A USA300 variant and other human-related methicillin-resistant *Staphylococcus aureus* strains infecting cats and dogs in France. *J. Antimicrob. Chemother.*, *67*(**2**), 326–9.

Hallin, M., Deplano, A., Denis, O., De Mendonça, R., De Ryck, R., and Struelens, M. J. (2007). Validation of pulsed-field gel electrophoresis and spa typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J. Clin. Microbiol.*, *45*(**1**), 127–133.

Hanselman, B. A., Kruth, S. A., Rousseau, J., and Weese, J. S. (2009). Coagulase positive staphylococcal colonization of humans and their household pets. *Can. Vet. J.*, *50*(**9**), 954–8.

Harris, S. R., Cartwright, E. J. P., Török, M. E., Holden, M. T. G., Brown, N. M., Ogilvy-Stuart, A. L., Ellington, M. J., Quail, M. A., Bentley, S. D., Parkhill, J., and Peacock, S. J. (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.*, *13*(**2**), 130–6.

Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J. A., Edgeworth, J. D., de Lencastre, H., Parkhill, J., Peacock, S. J., and Bentley, S. D. (2010). Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science (80-.).*, *327*(**5964**), 469–474.

Harrison, E. M., Paterson, G. K., Holden, M. T. G., Larsen, J., Stegger, M., Larsen, A. R., Petersen, A., Skov, R. L., Christensen, J. M., Bak Zeuthen, A., Heltberg, O., Harris, S. R., Zadoks, R. N., Parkhill, J., Peacock, S. J., and Holmes, M. A. (2013). Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol. Med.*, *5*(**4**), 509–515.

Harrison, E. M., Weinert, L. A., Holden, M. T. G., Welch, J. J., Wilson, K., Morgan, F. J. E., Harris, S. R., Loeffler, A., Boag, A. K., Peacock, S. J., Paterson, G. K., Waller, A. S., Parkhill, J., and Holmes, M. A. (2014). A shared population of Epidemic Methicillin-resistant *Staphylococcus aureus* 15 circulates in humans and companion animals. *MBio*, *5*(**3**), e00985–13.

Hartleib, J., Kohler, N., Dickinson, R. B., Chhatwal, G. S., Sixma, J. J., Hartford, O. M., Foster, T. J., Peters, G., Kehrel, B. E., and Herrmann, M. (2000). Protein A is the von Willebrand factor binding protein on *Staphylococcus aureus*. *Blood*, *96*(**6**), 2149–2156.

Hasman, H., Moodley, A., Guardabassi, L., Stegger, M., Skov, R. L., and Aarestrup, F. M. (2010). *Spa* type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. *Vet. Microbiol.*, *141*(**3-4**), 326–31.

Hata, E., Katsuda, K., Kobayashi, H., Uchida, I., Tanaka, K., and Eguchi, M. (2010). Genetic variation among *Staphylococcus aureus* strains from bovine milk and their relevance to methicillin-resistant isolates from humans. *J. Clin. Microbiol.*, *48*(**6**), 2130–2139.

Havaei, S. A., Vidovic, S., Tahmineh, N., Mohammad, K., Mohsen, K., Starnino, S., and Dillon, J.-A. R. (2011). Epidemic methicillin-susceptible *Staphylococcus aureus* lineages are the main cause of infections at an Iranian university hospital. *J. Clin. Microbiol.*, *49*(**11**), 3990–3.

He, M., Sebaihia, M., Lawley, T. D., Stabler, R. A., Dawson, L. F., Martin, M. J., Holt, K. E., Seth-Smith, H., Quail, M. A., Rance, R., and others. (2010). Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci.*, *107*(**16**), 7527.

Hébert, A., Sayasith, K., Sénéchal, S., Dubreuil, P., and Lagacé, J. (2000). Demonstration of intracellular *Staphylococcus aureus* in bovine mastitis alveolar cells and macrophages isolated from naturally infected cow milk. *FEMS Microbiol. Lett.*, *193*(**1**), 57–62.

Herold, B. C., Immergluck, L. C., Maranan, M. C., Lauderdale, D. S., Gaskin, R. E., Boyle-Vavra, S., Leitch, C. D., and Daum, R. S. (1998). Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. *Jama*, *279*(**8**), 593–598.

Herron-Olson, L., Fitzgerald, J. R., Musser, J. M., and Kapur, V. (2007). Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS One*, *2*(**10**), e1120.

Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., Roach, J. C., Kremer, K., Petrov, D. A., Feldman, M. W., and Gagneux, S. (2008). High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.*, *6*(**12**), e311.

Hidron, A. I., Low, C. E., Honig, E. G., and Blumberg, H. M. (2009). Emergence of community-acquired meticillin-resistant *Staphylococcus aureus* strain USA300 as a cause of necrotising community-onset pneumonia. *Lancet Infect. Dis.*, *9*(**6**), 384–92.

Highlander, S. K., Hultén, K. G., Qin, X., Jiang, H., Yerrapragada, S., Mason, E. O., Shang, Y., Williams, T. M., Fortunov, R. M., Liu, Y., Igboeli, O., Petrosino, J., Tirumalai, M., Uzman, A., Fox, G. E., Cardenas, A. M., Muzny, D. M., Hemphill, L., Ding, Y., Dugan, S., Blyth, P. R., Buhay, C. J., Dinh, H. H., Hawes, A. C., Holder, M., Kovar, C. L., Lee, S. L., Liu, W., Nazareth, L. V, Wang, Q., Zhou, J., Kaplan, S. L., and Weinstock, G. M. (2007). Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol.*, *7*(**1**), 99.

Hiller, N. L., Ahmed, A., Powell, E., Martin, D. P., Eutsey, R., Earl, J., Janto, B., Boissy, R. J., Hogg, J., Barbadora, K., Sampath, R., Lonergan, S., Post, J. C., Hu, F. Z., and Ehrlich, G. D. (2010). Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog*, *6*(**9**), e1001108.

Hillerton, J. E., and Berry, E. A. (2005). Treating mastitis in the cow - a tradition or an archaism. *J. Appl. Microbiol.*, *98*(**6**), 1250–1255.

Himsworth, C. G., Miller, R. R., Montoya, V., Hoang, L., Romney, M. G., Al-Rawahi, G. N., Kerr, T., Jardine, C. M., Patrick, D. M., Tang, P., and Weese, J. S. (2014). Carriage of methicillin-resistant *Staphylococcus aureus* by wild urban Norway rats (Rattus norvegicus). *PLoS One*, *9*(**2**), e87983.

Hiramatsu, K., Cui, L., Kuroda, M., and Ito, T. (2001). The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol.*, *9*(**10**), 486–493.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T., and Tenover, F. C. (1997). Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *J. Antimicrob. Chemother.*, *40*(**1**), 135–136.

Ho, P.-L., Lo, P.-Y., Chow, K.-H., Lau, E. H. Y., Lai, E. L., Cheng, V. C. C., and Kao, R. Y. (2010). Vancomycin MIC creep in MRSA isolates from 1997 to 2008 in a healthcare region in Hong Kong. *J. Infect.*, *60*(**2**), 140–145.

Holden, M. T. G., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P. J., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G., and Parkhill, J. (2004). Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. U. S. A.*, *101*(**26**), 9786–9791.

Holden, M. T. G., Hsu, L.-Y., Kurt, K., Weinert, L. A., Mather, A. E., Harris, S. R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., Skov, R., Westh, H., Zemlicková, H., Coombs, G., Kearns, A. M., Hill, R. L. R., Edgeworth, J., Gould, I., Gant, V., Cooke, J., Edwards, G. F., McAdam, P. R., Templeton, K. E., McCann, A., Zhou, Z., Castillo-Ramírez, S., Feil, E. J., Hudson, L. O., Enright, M. C., Balloux, F., Aanensen, D. M., Spratt, B. G., Fitzgerald, J. R., Parkhill, J., Achtman, M., Bentley, S. D., and Nübel, U. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.*, *23*(**4**), 653–64.

Holden, M. T. G., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., Batra, R., Parkhill, J., Bentley, S. D., and Edgeworth, J. D. (2009). Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of Methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *J. Bacteriol.*, *192*(**3**), 888–892.

Holt, D. C., Holden, M. T. G., Tong, S. Y. C., Castillo-Ramirez, S., Clarke, L., Quail, M. A., Currie, B. J., Parkhill, J., Bentley, S. D., Feil, E. J., and Giffard, P. M. (2011). A very early-branching *Staphylococcus aureus* lineage lacking the carotenoid pigment staphyloxanthin. *Genome Biol. Evol.*, *3*, 881–895.

Houston, P., Rowe, S. E., Pozzi, C., Waters, E. M., and O'Gara, J. P. (2011). Essential role for the major autolysin in the fibronectin-binding protein-

mediated *Staphylococcus aureus* biofilm phenotype. *Infect. Immun.*, *79*(**3**), 1153–65.

Howden, B. P., Seemann, T., Harrison, P. F., McEvoy, C. R., Stanton, J.-A. L., Rand, C. J., Mason, C. W., Jensen, S. O., Firth, N., Davies, J. K., Johnson, P. D. R., and Stinear, T. P. (2010). Complete genome sequence of *Staphylococcus aureus* strain JKD6008, an ST239 clone of methicillin-resistant *Staphylococcus aureus* with intermediate-level vancomycin resistance. *J. Bacteriol.*, *192*(**21**), 5848–9.

Huang, T.-W., Chen, F.-J., Miu, W.-C., Liao, T.-L., Lin, A.-C., Huang, I.-W., Wu, K.-M., Tsai, S.-F., Chen, Y.-T., and Lauderdale, T.-L. Y. (2012). Complete genome sequence of *Staphylococcus aureus* M013, a pvl-positive, ST59-SCC*mec* type V strain isolated in Taiwan. *J. Bacteriol.*, *194*(**5**), 1256–7.

Huijsdens, X. W., van Dijke, B. J., Spalburg, E., van Santen-Verheuvel, M. G., Heck, M. E. O. C., Pluister, G. N., Voss, A., Wannet, W. J. B., and de Neeling, A. J. (2006). Community-acquired MRSA and pig-farming. *Ann. Clin. Microbiol. Antimicrob.*, *5*, 26.

Huson, D. H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.*, *23*(**2**), 254 –267.

Ikawaty, R., Brouwer, E. C., Jansen, M. D., van Duijkeren, E., Mevius, D., Verhoef, J., and Fluit, A. C. (2009). Characterization of Dutch *Staphylococcus aureus* from bovine mastitis using a Multiple Locus Variable Number Tandem Repeat Analysis. *Vet. Microbiol.*, *136*(**3-4**), 277–284.

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., and McVean, G. (2012). *De novo* assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.*, *44*(**2**), 226–32.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., and Hiramatsu, K. (2001). Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *45*(**5**), 1323–1336.

Ito, T., Katayama, Y., and Hiramatsu, K. (1999). Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob. Agents Chemother.*, *43*(**6**), 1449–58.

Ito, T., Ma, X. X., Takeuchi, F., Okuma, K., Yuzawa, H., and Hiramatsu, K. (2004). Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. *Antimicrob. Agents Chemother.*, *48*(**7**), 2637–2651.

IWG-SCC. (2009). Classification of Staphylococcal Cassette Chromosome mec (SCC*mec*): Guidelines for Reporting Novel SCC*mec* Elements. *Antimicrob. Agents Chemother.*, *53*(**12**), 4961–4967.

Jackson, M. P., and Iandolo, J. J. (1986). Cloning and expression of the exfoliative toxin B gene from *Staphylococcus aureus*. *J. Bacteriol.*, *166*(**2**), 574–580.

Jakobsen, R. A., Heggebø, R., Sunde, E. B., and Skjervheim, M. (2011). *Staphylococcus aureus* and *Listeria monocytogenes* in Norwegian raw milk cheese production. *Food Microbiol.*, *28*(**3**), 492–6.

Jevons. (1961). Celbenin-resistant *Staphylococci*. *Br. Med. J.*, *1*(**5219**), 124–126.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., Wang, J., Liu, H., Yang, J., Yang, F., Zhang, X., Zhang, J., Yang, G., Wu, H., Qu, D., Dong, J., Sun, L., Xue, Y., Zhao, A., Gao, Y., Zhu, J., Kan, B., Ding, K., Chen, S., Cheng, H., Yao, Z., He, B., Chen, R., Ma, D., Qiang, B., Wen, Y., Hou, Y., and Yu, J. (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.*, *30*(**20**), 4432–41.

Jones, B. A., Grace, D., Kock, R., Alonso, S., Rushton, J., Said, M. Y., McKeever, D., Mutua, F., Young, J., McDermott, J., and Pfeiffer, D. U. (2013). Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(**21**), 8399–404.

Jørgensen, C. J., Cavaco, L. M., Hasman, H., Emborg, H.-D., and Guardabassi, L. (2007). Occurrence of CTX-M-1-producing *Escherichia coli* in pigs treated with ceftiofur. *J. Antimicrob. Chemother.*, *59*(**5**), 1040–2.

Jørgensen, H. J., Mørk, T., Caugant, D. A., Kearns, A., and Rørvik, L. M. (2005). Genetic variation among *Staphylococcus aureus* strains from Norwegian bulk milk. *Appl. Environ. Microbiol.*, *71*(**12**), 8352–8361.

Juhász-Kaszanyitzky, E., Jánosi, S., Somogyi, P., Dán, A., van der Graaf-van Bloois, L., van Duijkeren, E., and Wagenaar, J. A. (2007). MRSA transmission between cows and humans. *Emerg. Infect. Dis.*, *13*(**4**), 630–632.

Kaatz, G. W., Seo, S. M., and Ruble, C. A. (1993). Efflux-mediated fluoroquinolone resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *37*(**5**), 1086–1094.

Kadlec, K., and Schwarz, S. (2009). Novel ABC transporter gene, *vga(C)*, located on a multiresistance plasmid from a porcine methicillin-resistant *Staphylococcus aureus* ST398 strain. *Antimicrob. Agents Chemother.*, *53*(**8**), 3589–91.

Kampen, A. H., Tollersrud, T., and Lund, A. (2005). *Staphylococcus aureus* capsular polysaccharide types 5 and 8 reduce killing by bovine neutrophils in vitro. *Infect. Immun.*, *73*(**3**), 1578–83.

Kang, M., Ko, Y.-P., Liang, X., Ross, C. L., Liu, Q., Murray, B. E., and Höök, M. (2013). Collagen-binding microbial surface components recognizing adhesive

matrix molecule (MSCRAMM) of Gram-positive bacteria inhibit complement activation via the classical pathway. *J. Biol. Chem.*, *288*(**28**), 20520–31.

Kapral, F. A., and Miller, M. M. (1971). Product of *Staphylococcus aureus* Responsible for the Scalded-Skin Syndrome. *Infect. Immun.*, *4*(**5**), 541–545.

Karauzum, H., Adhikari, R. P., Sarwar, J., Devi, V. S., Abaandou, L., Haudenschild, C., Mahmoudieh, M., Boroun, A. R., Vu, H., Nguyen, T., Warfield, K. L., Shulenin, S., and Aman, M. J. (2013). Structurally designed attenuated subunit vaccines for *S. aureus* LukS-PV and LukF-PV confer protection in a mouse bacteraemia model. *PLoS One*, *8*(**6**), e65384.

Katayama, Y., Ito, T., and Hiramatsu, K. (2000). A new class of genetic element, Staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *44*(**6**), 1549–1555.

Katayama, Y., Ito, T., and Hiramatsu, K. (2001). Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: role of IS431-mediated *mecI* deletion in expression of resistance in *mecA*-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*. *Antimicrob. Agents Chemother.*, *45*(**7**), 1955.

Kechrid, A., Pérez-Vázquez, M., Smaoui, H., Hariga, D., Rodríguez-Baños, M., Vindel, A., Baquero, F., Cantón, R., and del Campo, R. (2011). Molecular analysis of community-acquired methicillin-susceptible and resistant *Staphylococcus aureus* isolates recovered from bacteraemic and osteomyelitis infections in children from Tunisia. *Clin. Microbiol. Infect.*, *17*(**7**), 1020–1026.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V, Liegeois, F., Loul, S., Ngole, E. M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M., and Hahn, B. H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, *313*(**5786**), 523–6.

Kelley, D. R., Schatz, M. C., and Salzberg, S. L. (2010). Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.*, *11*(**11**), R116.

Khan, S. A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid*, *53*(**2**), 126–36.

Khanna, T., Friendship, R., Dewey, C., and Weese, J. S. (2008). Methicillin resistant *Staphylococcus aureus* colonization in pigs and pig farmers. *Vet. Microbiol.*, *128*(**3-4**), 298–303.

King, M. D. (2006). Emergence of community-acquired methicillin-resistant *Staphylococcus aureus* USA 300 clone as the predominant cause of skin and soft-tissue infections. *Ann. Intern. Med.*, *144*(**5**), 309.

Klevens, R. M., Morrison, M. A., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L. H., Lynfield, R., Dumyati, G., Townes, J. M., and others. (2007). Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *Jama*, *298*(**15**), 1763.

Kluytmans, J., Van Belkum, A., and Verbrugh, H. (1997). Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin. Microbiol. Rev.*, *10*(**3**), 505–520.

Knight, G.M., Budd, E.L., Whitney, L., Thornley, A., Al-Ghusein, H., Planche, T., and Lindsay, J.A. (2012). Shift in dominant hosptial-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) clones over time. *J. Antimicrob. Chemother.*, *67*(**10**), 2514-2522.

Ko, K. S., Lee, J.-Y., Suh, J. Y., Oh, W. S., Peck, K. R., Lee, N. Y., and Song, J.-H. (2005). Distribution of major genotypes among methicillin-resistant *Staphylococcus aureus* clones in Asian countries. *J. Clin. Microbiol.*, *43*(**1**), 421–6.

Kobayashi, S. D., and DeLeo, F. R. (2013). *Staphylococcus aureus* protein A promotes immune suppression. *MBio*, *4*(**5**), e00764–13.

Komatsuzawa, H., Ohta, K., Sugai, M., Fujiwara, T., Glanzmann, P., Berger-BächiB, and Suginaka, H. (2000). Tn551-mediated insertional inactivation of the *fmtB* gene encoding a cell wall-associated protein abolishes methicillin resistance in *Staphylococcus aureus*. *J. Antimicrob. Chemother.*, *45*(**4**), 421–431.

Köser, C. U., Holden, M. T. G., Ellington, M. J., Cartwright, E. J. P., Brown, N. M., Ogilvy-Stuart, A. L., Hsu, L. Y., Chewapreecha, C., Croucher, N. J., Harris, S. R., Sanders, M., Enright, M. C., Dougan, G., Bentley, S. D., Parkhill, J., Fraser, L. J., Betley, J. R., Schulz-Trieglaff, O. B., Smith, G. P., and Peacock, S. J. (2012). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N. Engl. J. Med.*, *366*(**24**), 2267–75.

Kourbatova, E. V, Halvosa, J. S., King, M. D., Ray, S. M., White, N., and Blumberg, H. M. (2005). Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA 300 clone as a cause of health care-associated infections among patients with prosthetic joint infections. *Am. J. Infect. Control*, *33*(**7**), 385–91.

Krziwanek, K., Metz-Gercek, S., and Mittermayer, H. (2011). Trends in the occurrence of MRSA strains in Upper Austria from 2006 to 2009. *Clin. Microbiol. Infect.*, *17*(**6**), 920–923.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S.,

Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K. (2001). Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet*, *357*(**9264**), 1225–1240.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J. (2005). The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl. Acad. Sci. U. S. A.*, *102*(**14**), 5174–5179.

Lattar, S. M., Tuchscherr, L. P. N., Centrón, D., Becker, K., Predari, S. C., Buzzola, F. R., Robinson, D. A., and Sordelli, D. O. (2012). Molecular fingerprinting of *Staphylococcus aureus* isolated from patients with osteomyelitis in Argentina and clonal distribution of the cap5(8) genes and of other selected virulence genes. *Eur. J. Clin. Microbiol. Infect. Dis.*, *31*(**10**), 2559–66.

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Tsoi, H.-W., Wong, B. H. L., Wong, S. S. Y., Leung, S.-Y., Chan, K.-H., and Yuen, K.-Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.*, *102*(**39**), 14040–5.

Laurent, F., Chardon, H., Haenni, M., Bes, M., Reverdy, M.-E., Madec, J.-Y., Lagier, E., Vandenesch, F., and Tristan, A. (2012). MRSA harboring *mecA* variant gene *mecC*, France. *Emerg. Infect. Dis.*, *18*(**9**), 1465–7.

Le Loir, Y., Baron, F., and Gautier, M. (2003). *Staphylococcus aureus* and food poisoning. *Genet. Mol. Res.*, *2*(**1**), 63–76.

Le Maréchal, C., Hernandez, D., Schrenzel, J., Even, S., Berkova, N., Thiéry, R., Vautor, E., Fitzgerald, J. R., François, P., and Le Loir, Y. (2011). Genome sequences of two *Staphylococcus aureus* ovine strains that induce severe (strain O11) and mild (strain O46) mastitis. *J. Bacteriol.*, *193*(**9**), 2353–4.

Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian Phylogeography Finds Its Roots. *PLoS Comput Biol*, *5*(**9**), e1000520.

Lerat, E., and Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res.*, *33*(**10**), 3125 –3132.

Levin-Edens, E., Soge, O. O., No, D., Stiffarm, A., Scott Meschke, J., and Roberts, M. C. (2011). Methicillin-resistant *Staphylococcus aureus* from northwest marine and fresh water recreational beaches. *FEMS Microbiol. Ecol.*, *79*(**2**), 412–420.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(**14**), 1754 –1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(**16**), 2078–2079.

Li, K. S., Guan, Y., Wang, J., Smith, G. J. D., Xu, K. M., Duan, L., Rahardjo, A. P., Puthavathana, P., Buranathai, C., Nguyen, T. D., Estoepangestie, A. T. S., Chaisingh, A., Auewarakul, P., Long, H. T., Hanh, N. T. H., Webby, R. J., Poon, L. L. M., Chen, H., Shortridge, K. F., Yuen, K. Y., Webster, R. G., and Peiris, J. S. M. (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*, *430*(**6996**), 209–13.

Li, S., Skov, R. L., Han, X., Larsen, A. R., Larsen, J., Sørum, M., Wulf, M., Voss, A., Hiramatsu, K., and Ito, T. (2011). Novel types of staphylococcal cassette chromosome *mec* elements identified in clonal complex 398 methicillin-resistant *Staphylococcus aureus* strains. *Antimicrob. Agents Chemother.*, *55*(**6**), 3046–50.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., and Wang, L.-F. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science*, *310*(**5748**), 676–9.

Lin, Y., Barker, E., Kislow, J., Kaldhone, P., Stemper, M. E., Pantrangi, M., Moore, F. M., Hall, M., Fritsche, T. R., Novicki, T., Foley, S. L., and Shukla, S. K. (2011). Evidence of multiple virulence subtypes in nosocomial and community-associated MRSA genotypes in companion animals from the upper midwestern and northeastern United States. *Clin. Med. Res.*, *9*(**1**), 7–16.

Lindsay, J. A., and Holden, M. T. G. (2004). *Staphylococcus aureus*: superbug, super genome? *Trends Microbiol.*, *12*(**8**), 378–85.

Lindsay, J. A., and Holden, M. T. G. (2006). Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct. Integr. Genomics*, *6*(**3**), 186–201.

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D., and Hinds, J. (2006). Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *J. Bacteriol.*, *188*(**2**), 669–676.

Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N., and Novick, R. P. (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.*, *29*(**2**), 527–43.

Loeffler, A., Boag, A. K., Sung, J., Lindsay, J. A., Guardabassi, L., Dalsgaard, A., Smith, H., Stevens, K. B., and Lloyd, D. H. (2005). Prevalence of methicillin-

resistant *Staphylococcus aureus* among staff and pets in a small animal referral hospital in the UK. *J. Antimicrob. Chemother.*, *56*(**4**), 692–7.

Loeffler, A., Pfeiffer, D. U., Lindsay, J. A., Magalhães, R. J. S., and Lloyd, D. H. (2011). Prevalence of and risk factors for MRSA carriage in companion animals: a survey of dogs, cats and horses. *Epidemiol. Infect.*, *139*(**7**), 1019–28.

Loncaric, I., Kübber-Heiss, A., Posautz, A., Stalder, G. L., Hoffmann, D., Rosengarten, R., and Walzer, C. (2013). Characterization of methicillin-resistant Staphylococcus spp. carrying the *mecC* gene, isolated from wildlife. *J. Antimicrob. Chemother.*, *68*(**10**), 2222–5.

Lowder, B. V, Guinane, C. M., Zakour, N. L., Weinert, L. A., Conway-Morris, A., Cartwright, R. A., Simpson, A. J., Rambaut, A., Nübel, U., and Fitzgerald, J. R. (2009). Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.*, *106*(**46**), 19545–19550.

Lowy, F. (1998). *Staphylococcus aureus* infections. *N. Engl. J. Med.*, *339*(**8**), 520–532.

Lozano, C., Gómez-Sanz, E., Benito, D., Aspiroz, C., Zarazaga, M., and Torres, C. (2011). *Staphylococcus aureus* nasal carriage, virulence traits, antibiotic resistance mechanisms, and genetic lineages in healthy humans in Spain, with detection of CC398 and CC97 strains. *Int. J. Med. Microbiol.*, *301*(**6**), 500–505.

Luong, T. T., Ouyang, S., Bush, K., and Lee, C. Y. (2002). Type 1 capsule genes of *Staphylococcus aureus* are carried in a Staphylococcal cassette chromosome genetic element. *J. Bacteriol.*, *184*(**13**), 3623–3629.

Luzzago, C., Locatelli, C., Franco, A., Scaccabarozzi, L., Gualdi, V., Viganò, R., Sironi, G., Besozzi, M., Castiglioni, B., Lanfranchi, P., Cremonesi, P., and Battisti, A. (2014). Clonal diversity, virulence-associated genes and antimicrobial resistance profile of *Staphylococcus aureus* isolates from nasal cavities and soft tissue infections in wild ruminants in Italian Alps. *Vet. Microbiol.*, *170*(**1**), 157–161.

Ma, X. X., Ito, T., Tiensasitorn, C., Jamklang, M., Chongtrakool, P., Boyle-Vavra, S., Daum, R. S., and Hiramatsu, K. (2002). Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. *Antimicrob. Agents Chemother.*, *46*(**4**), 1147–52.

Maiques, E., Úbeda, C., Tormo, M. Á., Ferrer, M. D., Lasa, Í., Novick, R. P., and Penadés, J. R. (2007). Role of Staphylococcal Phage and SaPI Integrase in Intra- and Interspecies SaPI Transfer. *J. Bacteriol.*, *189*(**15**), 5608–5616.

Malachowa, N., and DeLeo, F. R. (2010). Mobile genetic elements of *Staphylococcus aureus*. *Cell. Mol. Life Sci.*, *67*(**18**), 3057–71.

Malik, S., Coombs, G. W., O'Brien, F. G., Peng, H., and Barton, M. D. (2006). Molecular typing of methicillin-resistant staphylococci isolated from cats and dogs. *J. Antimicrob. Chemother.*, *58*(**2**), 428–31.

Mangili, A., Bica, I., Snydman, D. R., and Hamer, D. H. (2005). Daptomycin-resistant, methicillin-resistant *Staphylococcus aureus* bacteraemia. *Clin. Infect. Dis.*, *40*(**7**), 1058–60.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, *26*(**19**), 2462–2463.

Martín, J. F., Barreiro, C., González-Lavado, E., and Barriuso, M. (2003). Ribosomal RNA and ribosomal proteins in corynebacteria. *J. Biotechnol.*, *104*(**1-3**), 41–53.

Marttinen, P., Hanage, W. P., Croucher, N. J., Connor, T. R., Harris, S. R., Bentley, S. D., and Corander, J. (2012). Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res.*, *40*(**1**), e6.

Mazmanian, S. K., Skaar, E. P., Gaspar, A. H., Humayun, M., Gornicki, P., Jelenska, J., Joachmiak, A., Missiakas, D. M., and Schneewind, O. (2003). Passage of heme-iron across the envelope of *Staphylococcus aureus*. *Science*, *299*(**5608**), 906–9.

McAdam, P. R., Holmes, A., Templeton, K. E., and Fitzgerald, J. R. (2011). Adaptive Evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One*, *6*(**9**), e24301.

McAdam, P. R., Templeton, K. E., Edwards, G. F., Holden, M. T. G., Feil, E. J., Aanensen, D. M., Bargawi, H. J. A., Spratt, B. G., Bentley, S. D., Parkhill, J., Enright, M. C., Holmes, A., Girvan, E. K., Godfrey, P. A., Feldgarden, M., Kearns, A. M., Rambaut, A., Robinson, D. A., and Fitzgerald, J. R. (2012). Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci.*, *109*(**23**), 9107–9112.

McCarthy, A. J., van Wamel, W., Vandendriessche, S., Larsen, J., Denis, O., Garcia-Graells, C., Uhlemann, A.-C., Lowy, F. D., Skov, R., and Lindsay, J. A. (2012). *Staphylococcus aureus* CC398 clade associated with human-to-human transmission. *Appl. Environ. Microbiol.*, *78*(**24**), 8845–8.

McCarthy, A. J., Witney, A. A., Gould, K. A., Moodley, A., Guardabassi, L., Voss, A., Denis, O., Broens, E. M., Hinds, J., and Lindsay, J. A. (2011). The Distribution of mobile genetic elements (MGEs) in MRSA CC398 is associated with both host and country. *Genome Biol. Evol.*, *3*(**0**), 1164–1174.

McCarthy, A., and Lindsay, J. (2010). Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen. *BMC Microbiol.*, *10*, 173.

McCarthy, A., and Lindsay, J. (2012). The distribution of plasmids that carry virulence and resistance genes in *Staphylococcus aureus* is lineage associated. *BMC Microbiol.*, *12*(**1**), 104.

McGavin, M. J., Arsic, B., and Nickerson, N. N. (2012). Evolutionary blueprint for host- and niche-adaptation in *Staphylococcus aureus* clonal complex CC30. *Front. Cell. Infect. Microbiol.*, *2*, 48.

McLaws, F. B., Larsen, A. R., Skov, R. L., Chopra, I., and O'Neill, A. J. (2011). Distribution of fusidic acid resistance determinants in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *55*(**3**), 1173 –1176.

McNamee, P. T., McCullagh, J. J., Thorp, B. H., Ball, H. J., Graham, D., McCullough, S. J., McConaghy, D., and Smyth, J. A. (1998). Study of leg weakness in two commercial broiler flocks. *Vet. Rec.*, *143*(**5**), 131 –135.

Medhus, A., Slettemeås, J. S., Marstein, L., Larssen, K. W., and Sunde, M. (2013). Methicillin-resistant *Staphylococcus aureus* with the novel *mecC* gene variant isolated from a cat suffering from chronic conjunctivitis. *J. Antimicrob. Chemother.*, *68*(**4**), 968–9.

Mediavilla, J. R., Chen, L., Mathema, B., and Kreiswirth, B. N. (2012). Global epidemiology of community-associated methicillin resistant *Staphylococcus aureus* (CA-MRSA). *Curr. Opin. Microbiol.*, *15*(**5**), 588–595.

Meemken, D., Blaha, T., Hotzel, H., Strommenger, B., Klein, G., Ehricht, R., Monecke, S., and Kehrenberg, C. (2013). Genotypic and phenotypic characterization of *Staphylococcus aureus* isolates from wild boars. *Appl. Environ. Microbiol.*, *79*(**5**), 1739–42.

Meemken, D., Blaha, T., Tegeler, R., Tenhagen, B.-A., Guerra, B., Hammerl, J. A., Hertwig, S., Käsbohrer, A., Appel, B., and Fetsch, A. (2010). Livestock associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) isolated from lesions of pigs at necropsy in Northwest Germany between 2004 and 2007. *Zoonoses Public Health*, *57*(**7-8**), 143–148.

Melles, D. C., Gorkink, R. F. J., Boelens, H. A. M., Snijders, S. V, Peeters, J. K., Moorhouse, M. J., van der Spek, P. J., van Leeuwen, W. B., Simons, G., Verbrugh, H. A., and van Belkum, A. (2004). Natural population dynamics and expansion of pathogenic clones of *Staphylococcus aureus*. *J. Clin. Invest.*, *114*(**12**), 1732–1740.

Mellmann, A., Harmsen, D., Cummings, C. A., Zentz, E. B., Leopold, S. R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S. F.,

Henkhaus, J. K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P. M., Moore, R. L., Guenther, S., Rothberg, J. M., and Karch, H. (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One*, *6*(**7**), e22751.

Merino, N., Toledo-Arana, A., Vergara-Irigaray, M., Valle, J., Solano, C., Calvo, E., Lopez, J. A., Foster, T. J., Penadés, J. R., and Lasa, I. (2009). Protein A-mediated multicellular behavior in *Staphylococcus aureus*. *J. Bacteriol.*, *191*(**3**), 832–43.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.*, *11*(**1**), 31–46.

Miles, H., Lesser, W., and Sears, P. (1992). The economic implications of bioengineered mastitis control. *J. Dairy Sci.*, *75*(**2**), 596–605.

Miller, L. G., Perdreau-Remington, F., Rieg, G., Mehdi, S., Perlroth, J., Bayer, A. S., Tang, A. W., Phung, T. O., and Spellberg, B. (2005). Necrotizing fasciitis caused by community-associated methicillin-resistant *Staphylococcus aureus* in Los Angeles. *N. Engl. J. Med.*, *352*(**14**), 1445.

Miller, R., Walker, A. S., Knox, K., Wyllie, D., Paul, J., Haworth, E., Mant, D., Peto, T., and Crook, D. W. (2010). "Feral"and "wild"-type methicillin-resistant *Staphylococcus aureus* in the United Kingdom. *Epidemiol. Infect.*, *138*(**05**), 655–665.

Monecke, S., Kuhnert, P., Hotzel, H., Slickers, P., and Ehricht, R. (2007). Microarray based study on virulence-associated genes and resistance determinants of *Staphylococcus aureus* isolates from cattle. *Vet. Microbiol.*, *125*(**1–2**), 128–140.

Monecke, S., Ruppelt, A., Wendlandt, S., Schwarz, S., Slickers, P., Ehricht, R., and Jäckel, S. C. de. (2013). Genotyping of *Staphylococcus aureus* isolates from diseased poultry. *Vet. Microbiol.*, *162*(**2-4**), 806–12.

Monecke, S., Skakni, L., Hasan, R., Ruppelt, A., Ghazal, S., Hakawi, A., Slickers, P., and Ehricht, R. (2012). Characterisation of MRSA strains isolated from patients in a hospital in Riyadh, Kingdom of Saudi Arabia. *BMC Microbiol.*, *12*(**1**), 146.

Monecke, S., Slickers, P., Hotzel, H., Richter-Huhn, G., Pohle, M., Weber, S., Witte, W., and Ehricht, R. (2006). Microarray-based characterisation of a Panton-Valentine leukocidin-positive community-acquired strain of methicillin-resistant *Staphylococcus aureus*. *Clin. Microbiol. Infect.*, *12*(**8**), 718–28.

Monk, I. R., Shah, I. M., Xu, M., Tan, M.-W., and Foster, T. J. (2012). Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *MBio*, *3*(**2**), e00277–11.

Montgomery, C. P., Boyle-Vavra, S., and Daum, R. S. (2009). The arginine catabolic mobile element is not associated with enhanced virulence in experimental invasive disease caused by the community-associated methicillin-resistant *Staphylococcus aureus* USA300 genetic background. *Infect. Immun.*, *77*(**7**), 2650–6.

Moodley, A., Espinosa-Gongora, C., Nielsen, S. S., McCarthy, A. J., Lindsay, J. A., and Guardabassi, L. (2012). Comparative host specificity of human- and pig-associated *Staphylococcus aureus* clonal lineages. *PLoS One*, *7*(**11**), e49344.

Moodley, A., Stegger, M., Bagcigil, A. F., Baptiste, K. E., Loeffler, A., Lloyd, D. H., Williams, N. J., Leonard, N., Abbott, Y., Skov, R., and Guardabassi, L. (2006). *Spa* typing of methicillin-resistant *Staphylococcus aureus* isolated from domestic animals and veterinary staff in the UK and Ireland. *J. Antimicrob. Chemother.*, *58*(**6**), 1118–1123.

Moran, G. J., Krishnadasan, A., Gorwitz, R. J., Fosheim, G. E., McDougal, L. K., Carey, R. B., and Talan, D. A. (2006). Methicillin-resistant *S. aureus* infections among patients in the emergency department. *N. Engl. J. Med.*, *355*(**7**), 666–74.

Moran, N. A., and Plague, G. R. (2004). Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.*, *14*(**6**), 627–633.

Morikawa, K., Takemura, A. J., Inose, Y., Tsai, M., Nguyen Thi, L. T., Ohta, T., and Msadek, T. (2012). Expression of a cryptic secondary sigma factor gene unveils natural competence for DNA transformation in *Staphylococcus aureus*. *PLoS Pathog.*, *8*(**11**), e1003003.

Morris, P. J., Johnson, W. R., Pisani, J., Bossart, G. D., Adams, J., Reif, J. S., and Fair, P. A. (2011). Isolation of culturable microorganisms from free-ranging bottlenose dolphins (*Tursiops truncatus*) from the southeastern United States. *Vet. Microbiol.*, *148*(**2-4**), 440–7.

Morrison-Rodriguez, S. M., Pacha, L. A., Patrick, J. E., and Jordan, N. N. (2010). Community-associated methicillin-resistant *Staphylococcus aureus* infections at an army training installation. *Epidemiol. Infect.*, *138* **05**), 721–729.

Mulders, M. N., Haenen, A. P. J., Geenen, P. L., Vesseur, P. C., Poldervaart, E. S., Bosch, T., Huijsdens, X. W., Hengeveld, P. D., Dam-Deisz, W. D. C., Graat, E. A. M., Mevius, D., Voss, A., and Van De Giessen, A. W. (2010). Prevalence of livestock-associated MRSA in broiler flocks and risk factors for slaughterhouse personnel in The Netherlands. *Epidemiol. Infect.*, *138*(**5**), 743–55.

Mulvey, M. R., MacDougall, L., Cholin, B., Horsman, G., Fidyk, M., and Woods, S. (2005). Community-associated methicillin-resistant *Staphylococcus aureus*, Canada. *Emerg. Infect. Dis.*, *11*(**6**), 844–850.

Murchan, S., Kaufmann, M. E., Deplano, A., de Ryck, R., Struelens, M., Zinn, C. E., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., El Solh, N., Cuny, C., Witte, W., Tassios, P. T., Legakis, N., van Leeuwen, W., van Belkum, A., Vindel, A., Laconcha, I., Garaizar, J., Haeggman, S., Olsson-Liljequist, B., Ransjo, U., Coombes, G., and Cookson, B. (2003). Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J. Clin. Microbiol.*, *41*(**4**), 1574–1585.

Musser, J. M., Schlievert, P. M., Chow, A. W., Ewan, P., Kreiswirth, B. N., Rosdahl, V. T., Naidu, A. S., Witte, W., and Selander, R. K. (1990). A single clone of *Staphylococcus aureus* causes the majority of cases of toxic shock syndrome. *Proc. Natl. Acad. Sci. U. S. A.*, *87*(**1**), 225–229.

Musser, J. M., and Selander, R. K. S. (1990). Genetic Analysis of natural populations of *Staphylococcus aureus*. In *Molecular Biology of the Staphylococci* (pp. 59–67). New York: VCH.

Muttaiyah, S., Coombs, G., Pandey, S., Reed, P., Ritchie, S., Lennon, D., and Roberts, S. (2010). Incidence, Risk Factors, and Outcomes of Panton-Valentine Leukocidin-Positive Methicillin-Susceptible *Staphylococcus aureus* Infections in Auckland, New Zealand. *J. Clin. Microbiol.*, *48*(**10**), 3470–3474.

Neela, V., Mohd Zafrul, A., Mariana, N. S., van Belkum, A., Liew, Y. K., and Rad, E. G. (2009). Prevalence of ST9 methicillin-resistant *Staphylococcus aureus* among pigs and pig handlers in Malaysia. *J. Clin. Microbiol.*, *47*(**12**), 4138–40.

Nickerson, S. C., Owens, W. E., and Boddie, R. L. (1995). Mastitis in dairy heifers: initial studies on prevalence and control. *J. Dairy Sci.*, *78*(**7**), 1607–18.

Noto, M. J., Kreiswirth, B. N., Monk, A. B., and Archer, G. L. (2008). Gene acquisition at the insertion site for SCCmec, the genomic island conferring methicillin resistance in *Staphylococcus aureus*. *J. Bacteriol.*, *190*(**4**), 1276–83.

Novick, R. (2003). Mobile genetic elements and bacterial toxinoses: the superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid*, *49*(**2**), 93–105.

Novick, R. P., Christie, G. E., and Penadés, J. R. (2010). The phage-related chromosomal islands of Gram-positive bacteria. *Nat. Rev. Microbiol.*, *8*(**8**), 541–551.

Nübel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., Žemličková, H., Leblois, R., Wirth, T., Jombart, T., Balloux, F., and Witte, W. (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.*, *6*(**4**), e1000855.

Nübel, U., Roumagnac, P., Feldkamp, M., Song, J.-H., Ko, K. S., Huang, Y.-C., Coombs, G., Ip, M., Westh, H., Skov, R., Struelens, M. J., Goering, R. V, Strommenger, B., Weller, A., Witte, W., and Achtman, M. (2008). Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.*, *105*(**37**), 14130–14135.

O'Mahony, R., Abbott, Y., Leonard, F. C., Markey, B. K., Quinn, P. J., Pollock, P. J., Fanning, S., and Rossney, A. S. (2005). Methicillin-resistant *Staphylococcus aureus* (MRSA) isolated from animals and veterinary personnel in Ireland. *Vet. Microbiol.*, *109*(**3-4**), 285–96.

O'Neill, E., Pozzi, C., Houston, P., Humphreys, H., Robinson, D. A., Loughman, A., Foster, T. J., and O'Gara, J. P. (2008). A novel *Staphylococcus aureus* biofilm phenotype mediated by the fibronectin-binding proteins, FnBPA and FnBPB. *J. Bacteriol.*, *190*(**11**), 3835–50.

O'Riordan, K., and Lee, J. C. (2004). *Staphylococcus aureus* Capsular Polysaccharides. *Clin. Microbiol. Rev.*, *17*(**1**), 218–234.

Okuma, K., Iwakawa, K., Turnidge, J. D., Grubb, W. B., Bell, J. M., O'Brien, F. G., Coombs, G. W., Pearman, J. W., Tenover, F. C., Kapi, M., Tiensasitorn, C., Ito, T., and Hiramatsu, K. (2002). Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. *J. Clin. Microbiol.*, *40*(**11**), 4289–4294.

Oliveira, D. C., Milheiriço, C., and de Lencastre, H. (2006). Redefining a structural variant of staphylococcal cassette chromosome *mec*, SCC*mec* type VI. *Antimicrob. Agents Chemother.*, *50*(**10**), 3457–3459.

Olsen, J. E., Christensen, H., and Aarestrup, F. M. (2006). Diversity and evolution of *blaZ* from *Staphylococcus aureus* and coagulase-negative staphylococci. *J. Antimicrob. Chemother.*, *57*(**3**), 450–60.

Olsen, S. J., Chang, H.-L., Cheung, T. Y.-Y., Tang, A. F.-Y., Fisk, T. L., Ooi, S. P.-L., Kuo, H.-W., Jiang, D. D.-S., Chen, K.-T., Lando, J., Hsu, K.-H., Chen, T.-J., and Dowell, S. F. (2003). Transmission of the severe acute respiratory syndrome on aircraft. *N. Engl. J. Med.*, *349*(**25**), 2416–22.

Ote, I., Taminiau, B., Duprez, J.-N., Dizier, I., and Mainil, J. G. (2011). Genotypic characterization by polymerase chain reaction of *Staphylococcus aureus* isolates associated with bovine mastitis. *Vet. Microbiol.*, *153*(**3**), 285–292.

Otter, J. A., and French, G. L. (2008). The emergence of community-associated methicillin-resistant *Staphylococcus aureus* at a London teaching hospital, 2000-2006. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.*, *14*(**7**), 670–676.

Otter, J. A., and French, G. L. (2010). Molecular epidemiology of community-associated meticillin-resistant *Staphylococcus aureus* in Europe. *Lancet Infect. Dis.*, *10*(**4**), 227–239.

Owens, W. E., Ray, C. H., Watts, J. L., and Yancey, R. J. (1997). Comparison of success of antibiotic therapy during lactation and results of antimicrobial susceptibility tests for bovine mastitis. *J. Dairy Sci.*, *80*(**2**), 313–7.

Parsonnet, J., Hansmann, M. A., Seymour, J. L., Delaney, M. L., DuBois, A. M., Modern, P. A., Jones, M. B., Wild, J. E., and Onderdonk, A. B. (2010). Persistence survey of Toxic Shock Syndrome toxin-1 producing *Staphylococcus aureus* and serum antibodies to this superantigen in five groups of menstruating women. *BMC Infect. Dis.*, *10*(**1**), 249.

Paterson, G. K., Larsen, A. R., Robb, A., Edwards, G. E., Pennycott, T. W., Foster, G., Mot, D., Hermans, K., Baert, K., Peacock, S. J., Parkhill, J., Zadoks, R. N., and Holmes, M. A. (2012). The newly described *mecA* homologue, *mecA*$_{\mathrm{LGA251}}$, is present in methicillin-resistant *Staphylococcus aureus* isolates from a diverse range of host species. *J. Antimicrob. Chemother.*, *67*(**12**), 2809–13.

Peacock, S. J., Day, N. P., Thomas, M. G., Berendt, A. R., and Foster, T. J. (2000). Clinical isolates of *Staphylococcus aureus* exhibit diversity in *fnb* genes and adhesion to human fibronectin. *J. Infect.*, *41*(**1**), 23–31.

Peck, K. R., Baek, J. Y., Song, J.-H., and Ko, K. S. (2009). Comparison of genotypes and enterotoxin genes between *Staphylococcus aureus* isolates from blood and nasal colonizers in a Korean hospital. *J. Korean Med. Sci.*, *24*(**4**), 585–591.

Pereira, U. P., Oliveira, D. G. S., Mesquita, L. R., Costa, G. M., and Pereira, L. J. (2011). Efficacy of *Staphylococcus aureus* vaccines for bovine mastitis: a systematic review. *Vet. Microbiol.*, *148*(**2-4**), 117–24.

Pérez-Losada, M., Browne, E. B., Madsen, A., Wirth, T., Viscidi, R. P., and Crandall, K. A. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.*, *6*(**2**), 97–112.

Petersen, A., Stegger, M., Heltberg, O., Christensen, J., Zeuthen, A., Knudsen, L. K., Urth, T., Sorum, M., Schouls, L., Larsen, J., Skov, R., and Larsen, A. R. (2013). Epidemiology of methicillin-resistant *Staphylococcus aureus* carrying the novel *mecC* gene in Denmark corroborates a zoonotic reservoir with transmission to humans. *Clin. Microbiol. Infect.*, *19*(**1**), 16–22.

Pishchany, G., McCoy, A. L., Torres, V. J., Krause, J. C., Crowe Jr., J. E., Fabry, M. E., and Skaar, E. P. (2010). Specificity for Human Hemoglobin Enhances *Staphylococcus aureus* Infection. *Cell Host Microbe*, *8*(**6**), 544–550.

Porrero, M. C., Mentaberre, G., Sánchez, S., Fernández-Llario, P., Gómez-Barrero, S., Navarro-Gonzalez, N., Serrano, E., Casas-Díaz, E., Marco, I., Fernández-Garayzabal, J.-F., Mateos, A., Vidal, D., Lavín, S., and Domínguez, L. (2013). Methicillin resistant *Staphylococcus aureus* (MRSA) carriage in different free-living wild animal species in Spain. *Vet. J.*, *198*(**1**), 127–30.

Porrero, M. C., Valverde, A., Fernández-Llario, P., Díez-Guerrier, A., Mateos, A., Lavín, S., Cantón, R., Fernández-Garayzabal, J.-F., and Domínguez, L. (2014). *Staphylococcus aureus* carrying *mecC* gene in animals and urban wastewater, Spain. *Emerg. Infect. Dis.*, *20*(**5**), 899–901.

Postma, B., Poppelier, M. J., van Galen, J. C., Prossnitz, E. R., van Strijp, J. A. G., de Haas, C. J. C., and van Kessel, K. P. M. (2004). Chemotaxis inhibitory protein of *Staphylococcus aureus* binds specifically to the C5a and formylated peptide receptor. *J. Immunol.*, *172*(**11**), 6994–7001.

Poutrel, B., Boutonnier, A., Sutra, L., and Fournier, J. M. (1988). Prevalence of capsular polysaccharide types 5 and 8 among *Staphylococcus aureus* isolates from cow, goat, and ewe milk. *J. Clin. Microbiol.*, *26*(**1**), 38–40.

Prasad, L. B., and Newbould, F. H. (1968). Inoculation of the bovine teat duct with Staph. Aureus: the relationship of teat duct length, milk yield and milking rate to development of intramammary infection. *Can. Vet. J.*, *9*(**5**), 107–115.

Price, L. B., Stegger, M., Hasman, H., Aziz, M., Larsen, J., Andersen, P. S., Pearson, T., Waters, A. E., Foster, J. T., Schupp, J., Gillece, J., Driebe, E., Liu, C. M., Springer, B., Zdovc, I., Battisti, A., Franco, A., Żmudzki, J., Schwarz, S., Butaye, P., Jouy, E., Pomba, C., Porrero, M. C., Ruimy, R., Smith, T. C., Robinson, D. A., Weese, J. S., Arriola, C. S., Yu, F., Laurent, F., Keim, P., Skov, R., and Aarestrup, F. M. (2012). *Staphylococcus aureus* CC398: Host Adaptation and Emergence of Methicillin Resistance in Livestock. *MBio*, *3*(**1**), e00305–11.

Pulliam, J. R. C., Epstein, J. H., Dushoff, J., Rahman, S. A., Bunning, M., Jamaluddin, A. A., Hyatt, A. D., Field, H. E., Dobson, A. P., and Daszak, P. (2012). Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface*, *9*(**66**), 89–101.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, *81*(**3**), 559–75.

Rabello, R. F., Moreira, B. M., Lopes, R. M. ., Teixeira, L. M., Riley, L. W., and Castro, A. C. . (2007). Multilocus sequence typing of *Staphylococcus aureus* isolates recovered from cows with mastitis in Brazilian dairy herds. *J. Med. Microbiol.*, *56*(**11**), 1505–1511.

Rahimpour, R., Mitchell, G., Khandaker, M. H., Kong, C., Singh, B., Xu, L., Ochi, A., Feldman, R. D., Pickering, J. G., Gill, B. M., and Kelvin, D. J. (1999). Bacterial superantigens induce down-modulation of CC chemokine responsiveness in human monocytes via an alternative chemokine ligand-independent mechanism. *J. Immunol. (Baltimore, Md. 1950)*, *162*(**4**), 2299–2307.

Rainard, P., Corrales, J.-C., Barrio, M. B., Cochard, T., and Poutrel, B. (2003). Leucotoxic Activities of *Staphylococcus aureus* Strains Isolated from Cows, Ewes, and Goats with Mastitis: Importance of LukM/LukF'-PV Leukotoxin. *Clin. Vaccine Immunol.*, *10*(**2**), 272–277.

Reinoso, E. B., El-Sayed, A., Lämmler, C., Bogni, C., and Zschöck, M. (2008). Genotyping of *Staphylococcus aureus* isolated from humans, bovine subclinical mastitis and food samples in Argentina. *Microbiol. Res.*, *163*(**3**), 314–322.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., and Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*, *25*(**16**), 2071–2073.

Rivero-Pérez, B., Alcoba-Flórez, J., and Méndez-Álvarez, S. (2012). Genetic diversity of community-associated methicillin-resistant *Staphylococcus aureus* isolated from Tenerife Island, Spain. *Infect. Genet. Evol.*, *12*(**3**), 586–90.

Roberts, G. A., Houston, P. J., White, J. H., Chen, K., Stephanou, A. S., Cooper, L. P., Dryden, D. T. F., and Lindsay, J. A. (2013). Impact of target site distribution for Type I restriction enzymes on the evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) populations. *Nucleic Acids Res.*, *41*(**15**), 7472–84.

Roberts, M. C., Soge, O. O., Horst, J. A., Ly, K. A., and Milgrom, P. (2011). Methicillin-resistant *Staphylococcus aureus* from dental school clinic surfaces and students. *Am. J. Infect. Control*, *39*(**8**), 628–632.

Roberts, M. C., Soge, O. O., No, D., Helgeson, S. E., and Meschke, J. S. (2011b). Characterization of methicillin-resistant *Staphylococcus aureus* isolated from public surfaces on a university campus, student homes and local community. *J. Appl. Microbiol.*, *110*(**6**), 1531–1537.

Robinson, D. A., and Enright, M. C. (2003). Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *47*(**12**), 3926–3934.

Robinson, D. A., and Enright, M. C. (2004). Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J. Bacteriol.*, *186*(**4**), 1060–1064.

Robinson, D. A., Kearns, A. M., Holmes, A., Morrison, D., Grundmann, H., Edwards, G., O'Brien, F. G., Tenover, F. C., McDougal, L. K., Monk, A. B.,

and Enright, M. C. (2005). Re-emergence of early pandemic *Staphylococcus aureus* as a community-acquired meticillin-resistant clone. *Lancet, 365*(**9466**), 1256–8.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nat. Biotechnol., 29*(**1**), 24–6.

Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N. J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., Xi, F., Li, S., Li, Y., Zhang, Z., Yang, X., Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., Kobbe, R., Loos, S., Oh, J., Yang, L., Danchin, A., Gao, G. F., Song, Y., Li, Y., Yang, H., Wang, J., Xu, J., Pallen, M. J., Wang, J., Aepfelbacher, M., and Yang, R. (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N. Engl. J. Med., 365*(**8**), 718–24.

Rolo, J., Miragaia, M., Turlej-Rogacka, A., Empel, J., Bouchami, O., Faria, N. A., Tavares, A., Hryniewicz, W., Fluit, A. C., de Lencastre, H., and and the CONCORD Working Group. (2012). High genetic diversity among community-associated *Staphylococcus aureus* in Europe: Results from a multicenter study. *PLoS One, 7*(**4**), e34768.

Rooijakkers, S. H. M., Ruyken, M., Roos, A., Daha, M. R., Presanis, J. S., Sim, R. B., van Wamel, W. J. B., van Kessel, K. P. M., and van Strijp, J. A. G. (2005). Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nat. Immunol., 6*(**9**), 920–927.

Rooijakkers, S. H. M., van Wamel, W. J. B., Ruyken, M., van Kessel, K. P. M., and van Strijp, J. A. G. (2005b). Anti-opsonic properties of staphylokinase. *Microbes Infect., 7*(**3**), 476–484.

Ruimy, R., Angebault, C., Djossou, F., Dupont, C., Epelboin, L., Jarraud, S., Armand Lefevre, L., Bes, M., Elena Lixandru, B., Bertine, M., El Miniai, A., Renard, M., Marc Bettinger, R., Lescat, M., Clermont, O., Peroz, G., Lina, G., Tavakol, M., Vandenesch, F., van Belkum, A., Rousset, F., and Andremont, A. (2010). Are host genetics the predominant determinant of persistent nasal *Staphylococcus aureus* carriage in humans? *J. Infect. Dis., 202*(**6**), 924 –934.

Ruimy, R., Armand-Lefevre, L., Barbier, F., Ruppé, E., Cocojaru, R., Mesli, Y., Maiga, A., Benkalfat, M., Benchouk, S., Hassaine, H., Dufourcq, J.-B., Nareth, C., Sarthou, J.-L., Andremont, A., and Feil, E. J. (2009). Comparisons between geographically diverse samples of carried *Staphylococcus aureus*. *J. Bacteriol., 191*(**18**), 5577 –5583.

Ruzin, A., Lindsay, J., and Novick, R. P. (2001). Molecular genetics of SaPI1 - a mobile pathogenicity island in *Staphylococcus aureus*. *Mol. Microbiol., 41*(**2**), 365–377.

Sabat, A., Krzyszton-Russjan, J., Strzalka, W., Filipek, R., Kosowska, K., Hryniewicz, W., Travis, J., and Potempa, J. (2003). New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. *J. Clin. Microbiol.*, *41*(**4**), 1801–1804.

Saiman, L., O'Keefe, M., Graham, P. L., Wu, F., Saïd-Salim, B., Kreiswirth, B., LaSala, A., Schlievert, P. M., and Della-Latta, P. (2003). Hospital transmission of community-acquired methicillin-resistant *Staphylococcus aureus* among postpartum women. *Clin. Infect. Dis.*, *37*(**10**), 1313–9.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A., and Moreillon, P. (2011). *Staphylococcus aureus* Host Range and Human-Bovine Host Shift. *Appl. Environ. Microbiol.*, *77*(**17**), 5908–5915.

Sam, I.-C., Kahar-Bador, M., Chan, Y.-F., Loong, S.-K., and Mohd Nor Ghazali, F. (2008). Multisensitive community-acquired methicillin-resistant *Staphylococcus aureus* infections in Malaysia. *Diagn. Microbiol. Infect. Dis.*, *62*(**4**), 437–439.

Sasaki, T., Tsubakishita, S., Tanaka, Y., Ohtsuka, M., Hongo, I., Fukata, T., Kabeya, H., Maruyama, S., and Hiramatsu, K. (2012). Population genetic structures of *Staphylococcus aureus* isolates from cats and dogs in Japan. *J. Clin. Microbiol.*, *50*(**6**), 2152–2155.

Schaumburg, F., Alabi, A. S., Köck, R., Mellmann, A., Kremsner, P. G., Boesch, C., Becker, K., Leendertz, F. H., and Peters, G. (2012). Highly divergent *Staphylococcus aureus* isolates from African non-human primates. *Environ. Microbiol. Rep.*, *4*(**1**), 141–6.

Schijffelen, M., Boel, C. E., van Strijp, J., and Fluit, A. (2010). Whole genome analysis of a livestock-associated methicillin-resistant *Staphylococcus aureus* ST398 isolate from a case of human endocarditis. *BMC Genomics*, *11*(**1**), 376.

Schroeder, K., Jularic, M., Horsburgh, S. M., Hirschhausen, N., Neumann, C., Bertling, A., Schulte, A., Foster, S., Kehrel, B. E., Peters, G., and Heilmann, C. (2009). Molecular characterization of a novel *Staphylococcus aureus* surface protein (SasC) involved in cell aggregation and biofilm accumulation. *PLoS One*, *4*(**10**), e7567.

Schuenck, R. P., Nouér, S. A., Winter, C. de O., Cavalcante, F. S., Scotti, T. D., Ferreira, A. L. P., Giambiagi-de Marval, M., and dos Santos, K. R. N. (2009). Polyclonal presence of non-multiresistant methicillin-resistant *Staphylococcus aureus* isolates carrying SCC*mec* IV in health care-associated infections in a hospital in Rio de Janeiro, Brazil. *Diagn. Microbiol. Infect. Dis.*, *64*(**4**), 434–441.

Schukken, Y. H., Leslie, K. E., Barnum, D. A., Mallard, B. A., Lumsden, J. H., Dick, P. C., Vessie, G. H., and Kehrli, M. E. (1999). Experimental *Staphylococcus*

aureus intramammary challenge in late lactation dairy cows: Quarter and cow effects determining the probability of infection. *J. Dairy Sci.*, *82*(**11**), 2393–2401.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics.*, *30*(**14**), 2068-2069.

Shearer, J. E. S., Wireman, J., Hostetler, J., Forberger, H., Borman, J., Gill, J., Sanchez, S., Mankin, A., Lamarre, J., Lindsay, J. A., Bayles, K., Nicholson, A., O'Brien, F., Jensen, S. O., Firth, N., Skurray, R. A., and Summers, A. O. (2011). Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. *G3 (Bethesda).*, *1*(**7**), 581–91.

Shepheard, M. A., Fleming, V. M., Connor, T. R., Corander, J., Feil, E. J., Fraser, C., and Hanage, W. P. (2013). Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset. *PLoS One*, *8*(**5**), e62369.

Shkreta, L., Talbot, B. G., Diarra, M. S., and Lacasse, P. (2004). Immune responses to a DNA/protein vaccination strategy against *Staphylococcus aureus* induced mastitis in dairy cows. *Vaccine*, *23*(**1**), 114–26.

Shopsin, B., Gomez, M., Montgomery, S. O., Smith, D. H., Waddington, M., Dodge, D. E., Bost, D. A., Riehman, M., Naidich, S., and Kreiswirth, B. N. (1999). Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J. Clin. Microbiol.*, *37*(**11**), 3556–3563.

Shore, A. C., Deasy, E. C., Slickers, P., Brennan, G., O'Connell, B., Monecke, S., Ehricht, R., and Coleman, D. C. (2011). Detection of Staphylococcal cassette chromosome *mec* type XI carrying highly divergent *mecA*, *mecI*, *mecR1*, *blaZ*, and *ccr* genes in human clinical isolates of clonal complex 130 Methicillin-Resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *55*(**8**), 3765–3773.

Shore, A. C., Rossney, A. S., Brennan, O. M., Kinnevey, P. M., Humphreys, H., Sullivan, D. J., Goering, R. V., Ehricht, R., Monecke, S., and Coleman, D. C. (2011b). Characterization of a novel arginine catabolic mobile element (ACME) and Staphylococcal chromosomal cassette *mec* composite island with significant homology to *Staphylococcus epidermidis* ACME type II in methicillin-resistant *Staphylococcus aureus* genotype ST22-MRSA-IV. *Antimicrob. Agents Chemother.*, *55*(**5**), 1896–1905.

Siboo, I. R., Chambers, H. F., and Sullam, P. M. (2005). Role of SraP, a Serine-Rich Surface Protein of *Staphylococcus aureus*, in Binding to Human Platelets. *Infect. Immun.*, *73*(**4**), 2273–2280.

Smith, E. M., Green, L. E., Medley, G. F., Bird, H. E., Fox, L. K., Schukken, Y. H., Kruze, J. V, Bradley, A. J., Zadoks, R. N., and Dowson, C. G. (2005).

Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. *J. Clin. Microbiol.*, *43*(**9**), 4737–4743.

Smith, E. M., Needs, P. F., Manley, G., and Green, L. E. (2013). Global distribution and diversity of ovine-associated *Staphylococcus aureus*. *Infect. Genet. Evol.*, *22*, 208–215.

Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S. M., Guan, Y., and Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, *459*(**7250**), 1122–5.

Smith, P. G. (2003). Bovine spongiform encephalopathy (BSE) and its epidemiology. *Br. Med. Bull.*, *66*(**1**), 185–198.

Smyth, D. S., Feil, E. J., Meaney, W. J., Hartigan, P. J., Tollersrud, T., Fitzgerald, J. R., Enright, M. C., and Smyth, C. J. (2009). Molecular genetic typing reveals further insights into the diversity of animal-associated *Staphylococcus aureus*. *J. Med. Microbiol.*, *58*(**Pt 10**), 1343–1353.

Smyth, D. S., McDougal, L. K., Gran, F. W., Manoharan, A., Enright, M. C., Song, J.-H., de Lencastre, H., and Robinson, D. A. (2010). Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. *PLoS One*, *5*(**1**), e8582.

Snitkin, E. S., Zelazny, A. M., Montero, C. I., Stock, F., Mijares, L., NISC Comparative Sequence Program, Murray, P. R., and Segre, J. A. (2011). Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proc. Natl. Acad. Sci.*, *108*(**33**), 13758 –13763.

Sol, J., Sampimon, O. C., Barkema, H. W., and Schukken, Y. H. (2000). Factors associated with cure after therapy of clinical mastitis caused by *Staphylococcus aureus*. *J. Dairy Sci.*, *83*(**2**), 278–284.

Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. ., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y. Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y., Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B., He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J., Spiga, O., Guo, Z. M., Pan, H. ., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y. X., Wu, C. I., and Zhao, G. P. (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. U. S. A.*, *102*(**7**), 2430–5.

Song, J. H., Hsueh, P. R., Chung, D. R., Ko, K. S., Kang, C. I., Peck, K. R., Yeom, J. S., Kim, S. W., Chang, H. H., Kim, Y. S., Jung, S. I., Son, J. S., So, T. M.,

Lalitha, M. K., Yang, Y., Huang, S. G., Wang, H., Lu, Q., Carlos, C. C., Perera, J. A., Chiu, C. H., Liu, J. W., Chongthaleong, A., Thamlikitkul, V., and Van, P. H. (2011). Spread of methicillin-resistant *Staphylococcus aureus* between the community and the hospitals in Asian countries: an ANSORP study. *J. Antimicrob. Chemother.*, *66*(**5**), 1061–9.

Spoor L.E., McAdam P.R., Weinert L.A., Rambaut A., Hasman H., Aarestrup F.M., Kearns A.M., Larsen A.R., Skov R.L., Fitzgerald J.R. (2013).). Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *mBio* 4(**4**), e00356-13.

Spratt, B. G., Hanage, W. P., and Feil, E. J. (2001). The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr. Opin. Microbiol.*, *4*(**5**), 602–606.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(**21**), 2688–2690.

Stam-Bolink, E. M., Mithoe, D., Baas, W. H., Arends, J. P., and Möller, A. V. M. (2007). Spread of a methicillin-resistant *Staphylococcus aureus* ST80 strain in the community of the northern Netherlands. *Eur. J. Clin. Microbiol. Infect. Dis.*, *26*(**10**), 723–7.

Stead, W. W., Eisenach, K. D., Cave, M. D., Beggs, M. L., Templeton, G. L., Thoen, C. O., and Bates, J. H. (1995). When did *Mycobacterium tuberculosis* infection first occur in the New World? An important question with public health implications. *Am. J. Respir. Crit. Care Med.*, *151*(**4**), 1267–8.

Stegger, M., Andersen, P. S., Kearns, A., Pichon, B., Holmes, M. A., Edwards, G., Laurent, F., Teale, C., Skov, R., and Larsen, A. R. (2012). Rapid detection, differentiation and typing of methicillin-resistant *Staphylococcus aureus* harbouring either *mecA* or the new *mecA* homologue $mecA_{(LGA251).}$ *Clin. Microbiol. Infect.*, *18*(**4**), 395–400.

Stegger, M., Driebe, E. M., Roe, C., Lemmer, D., Bowers, J. R., Engelthaler, D. M., Keim, P., and Andersen, P. S. (2013). Genome sequence of *Staphylococcus aureus* strain CA-347, a USA600 methicillin-resistant isolate. *Genome Announc.*, *1*(**4**), e00517–13.

Stegger, M., Price, L. B., Larsen, A. R., Gillece, J. D., Waters, A. E., Skov, R., and Andersen, P. S. (2012). Genome sequence of *Staphylococcus aureus* strain 11819-97, an ST80-IV European community-acquired methicillin-resistant isolate. *J. Bacteriol.*, *194*(**6**), 1625–6.

Strommenger, B., Braulke, C., Pasemann, B., Schmidt, C., and Witte, W. (2007). Multiplex PCR for Rapid Detection of *Staphylococcus aureus* Isolates

Suspected to Represent Community-Acquired Strains. *J. Clin. Microbiol.*, *46*(**2**), 582–587.

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics*, *27*(**7**), 1009–10.

Sung, J. M. ., Lloyd, D. H., and Lindsay, J. A. (2008). *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology*, *154*(**7**), 1949.

Sutra, L., and Poutrel, B. (1994). Virulence factors involved in the pathogenesis of bovine intramammary infections due to *Staphylococcus aureus*. *J. Med. Microbiol.*, *40*(**2**), 79–89.

Swofford, D. (2003). PAUP*: phylogenetic analysis using parsimony (*and other methods), *version 4.0b10* (Sinauer Associates, Inc., Sunderland, MA.).

Takahashi, H., Kikuchi, T., Shoji, S., Fujimura, S., Lutfor, A. B., Tokue, Y., Nukiwa, T., and Watanabe, A. (1998). Characterization of *gyrA*, *gyrB*, *grlA* and *grlB* mutations in fluoroquinolone-resistant clinical isolates of *Staphylococcus aureus*. *J. Antimicrob. Chemother.*, *41*(**1**), 49–57.

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M., Ankai, A., Baba, S., Fukui, S., Lee, J. C., and Hiramatsu, K. (2005). Whole-genome sequencing of *Staphylococcus haemolyticus* uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *J. Bacteriol.*, *187*(**21**), 7292–308.

Takeuchi, S., Matsunaga, K., Inubushi, S., Higuchi, H., Imaizumi, K., and Kaidoh, T. (2002). Structural gene and strain specificity of a novel cysteine protease produced by *Staphylococcus aureus* isolated from a diseased chicken. *Vet. Microbiol.*, *89*(**2-3**), 201–210.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, *24*(**8**), 1596–1599.

Tawee Chotpitayasunond, Ungchusak, K., Hanshaoworakul, W., Chunsuthiwat, S., Sawanpanyalert, P., Kijphati, R., Sorasak Lochindarat, P. S., Suwan, P., Osotthanakorn, Y., Anantasetagoon, T., Ka, S., and Dowell, S. F. (2005). Human Disease from Influenza A (H5N1), Thailand, 2004. *Emerg. Infect. Dis.*, *11*(**2**), 201–209.

Teixeira, L., Resende, C., Ormonde, L., Rosenbaum, R., Figueiredo, A., de Lencastre, H., and Tomasz, A. (1995). Geographic spread of epidemic multiresistant *Staphylococcus aureus* clone in Brazil. *J. Clin. Microbiol.*, *33*(**9**), 2400–2404.

Tenover, F. C., and Goering, R. V. (2009). Methicillin-resistant *Staphylococcus aureus* strain USA300: origin and epidemiology. *J. Antimicrob. Chemother.*, *64*(**3**), 441–6.

Thammavongsa, V., Kern, J. W., Missiakas, D. M., and Schneewind, O. (2009). *Staphylococcus aureus* synthesizes adenosine to escape host immune responses. *J. Exp. Med.*, *206*(**11**), 2417 –2427.

Thomas, J. C., Godfrey, P. A., Feldgarden, M., and Robinson, D. A. (2012). Draft genome sequences of *Staphylococcus aureus* sequence type 34 (ST34) and ST42 hybrids. *J. Bacteriol.*, *194*(**10**), 2740–1.

Thurlow, L. R., Joshi, G. S., Clark, J. R., Spontak, J. S., Neely, C. J., Maile, R., and Richardson, A. R. (2013). Functional Modularity of the Arginine Catabolic Mobile Element Contributes to the Success of USA300 Methicillin-Resistant *Staphylococcus aureus*. *Cell Host Microbe*, *13*(**1**), 100–7.

Tock, M. R., and Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, *8*(**4**), 466–72.

Tokajian, S. T., Khalil, P. A., Jabbour, D., Rizk, M., Farah, M. J., Hashwa, F. A., and Araj, G. F. (2010). Molecular characterization of *Staphylococcus aureus* in Lebanon. *Epidemiol. Infect.*, *138*(**5**), 707–712.

Tollersrud, T., Kenny, K., Reitz, A. J. . J., and Lee, J. C. (2000). Genetic and serologic evaluation of capsule production by bovine mammary isolates of *Staphylococcus aureus* and other *Staphylococcus* spp. from Europe and the United States. *J. Clin. Microbiol.*, *38*(**8**), 2998–3003.

Tormo-Más, M. A., Mir, I., Shrestha, A., Tallent, S. M., Campoy, S., Lasa, I., Barbé, J., Novick, R. P., Christie, G. E., and Penadés, J. R. (2010). Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature*, *465*(**7299**), 779–82.

Torres, V. J., Pishchany, G., Humayun, M., Schneewind, O., and Skaar, E. P. (2006). *Staphylococcus aureus* IsdB is a hemoglobin receptor required for heme iron utilization. *J. Bacteriol.*, *188*(**24**), 8421–9.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M. El, Frapy, E., Garry, L., Ghigo, J. M., Gilles, A. M., Johnson, J., Le Bouguénec, C., Lescat, M., Mangenot, S., Martinez-Jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M. A., Pichon, C., Rouy, Z., Ruf, C. Saint, Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E. P. C., and Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet*, *5*(**1**), e1000344.

Tran, T. H., Nguyen, T. L., Nguyen, T. D., Luong, T. S., Pham, P. M., Nguyen, van V. C., Pham, T. S., Vo, C. D., Le, T. Q. M., Ngo, T. T., Dao, B. K., Le, P. P., Nguyen, T. T., Hoang, T. L., Cao, V. T., Le, T. G., Nguyen, D. T., Le, H. N., Nguyen, K. T., Le, H. S., Le, V. T., Christiane, D., Tran, T. T., Menno, de J., Schultsz, C., Cheng, P., Lim, W., Horby, P., and Farrar, J. (2004). Avian influenza A (H5N1) in 10 patients in Vietnam. *N. Engl. J. Med.*, *350*(**12**), 1179–88.

Tristan, A., Bes, M., Meugnier, H., Lina, G., Bozdogan, B., Courvalin, P., Reverdy, M.-E., Enright, M. C., Vandenesch, F., and Etienne, J. (2007). Global distribution of Panton-Valentine leukocidin--positive methicillin-resistant *Staphylococcus aureus*, 2006. *Emerg. Infect. Dis.*, *13*(**4**), 594–600.

Tsiodras, S., Gold, H. S., Sakoulas, G., Eliopoulos, G. M., Wennersten, C., Venkataraman, L., Moellering, R. C., and Ferraro, M. J. (2001). Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. *Lancet*, *358*(**9277**), 207–8.

Úbeda, C., Maiques, E., Knecht, E., Lasa, Í., Novick, R. P., and Penadés, J. R. (2005). Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.*, *56*(**3**), 836–844.

Úbeda, C., Tormo, M. Á., Cucarella, C., Trotonda, P., Foster, T. J., Lasa, Í., and Penadés, J. R. (2003). Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol. Microbiol.*, *49*(**1**), 193–210.

Udo, E. E., Aly, N. Y. A., Sarkhoo, E., Al-Sawan, R., and Al-Asar, A.-S. M. (2011). Detection and characterization of an ST97-SCC*mec*-V community-associated meticillin-resistant *Staphylococcus aureus* clone in a neonatal intensive care unit and special care baby unit. *J. Med. Microbiol.*, *60*(**5**), 600 –604.

Uhlemann, A.-C., Porcella, S. F., Trivedi, S., Sullivan, S. B., Hafer, C., Kennedy, A. D., Barbian, K. D., McCarthy, A. J., Street, C., Hirschberg, D. L., Lipkin, W. I., Lindsay, J. A., DeLeo, F. R., and Lowy, F. D. (2012). Identification of a Highly Transmissible Animal-Independent *Staphylococcus aureus* ST398 Clone with Distinct Genomic and Cell Adhesion Properties. *MBio*, *3*(**2**), e00027–12.

Utsui, Y., and Yokota, T. (1985). Role of an altered penicillin-binding protein in methicillin- and cephem-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *28*(**3**), 397–403.

Van Cleef, B. A. G. L., Broens, E. M., Voss, A., Huijsdens, X. W., Züchner, L., Van Benthem, B. H. B., Kluytmans, J. A. J. W., Mulders, M. N., and Van De Giessen, A. W. (2010). High prevalence of nasal MRSA carriage in slaughterhouse workers in contact with live pigs in The Netherlands. *Epidemiol. Infect.*, *138*(**5**), 756–63.

Van de Giessen, A. W., van Santen-Verheuvel, M. G., Hengeveld, P. D., Bosch, T., Broens, E. M., and Reusken, C. B. E. M. (2009). Occurrence of methicillin-resistant *Staphylococcus aureus* in rats living on pig farms. *Prev. Vet. Med.*, *91*(**2–4**), 270–273.

Van Den Broek, I. V. F., Van Cleef, B. A. G. L., Haenen, A., Broens, E. M., Van Der Wolf, P. J., Van Den Broek, M. J. M., Huijsdens, X. W., Kluytmans, J. A. J. W., Van De Giessen, A. W., and Tiemersma, E. W. (2009). Methicillin-resistant *Staphylococcus aureus* in people living and working in pig farms. *Epidemiol. Infect.*, *137*(**5**), 700–8.

Van der Bij, A. K., and Pitout, J. D. D. (2012). The role of international travel in the worldwide spread of multiresistant Enterobacteriaceae. *J. Antimicrob. Chemother.*, *67*(**9**), 2090–100.

Van Duijkeren, E., Ikawaty, R., Broekhuizen-Stins, M. J., Jansen, M. D., Spalburg, E. C., de Neeling, A. J., Allaart, J. G., van Nes, A., Wagenaar, J. A., and Fluit, A. C. (2008). Transmission of methicillin-resistant *Staphylococcus aureus* strains between different kinds of pig farms. *Vet. Microbiol.*, *126*(**4**), 383–389.

Van Duijkeren, E., Moleman, M., Sloet van Oldruitenborgh-Oosterbaan, M. M., Multem, J., Troelstra, A., Fluit, A. C., van Wamel, W. J. B., Houwers, D. J., de Neeling, A. J., and Wagenaar, J. A. (2010). Methicillin-resistant *Staphylococcus aureus* in horses and horse personnel: an investigation of several outbreaks. *Vet. Microbiol.*, *141*(**1-2**), 96–102.

Van Duijkeren, E., Wolfhagen, M. J. H. M., Box, A. T. A., Heck, M. E. O. C., Wannet, W. J. B., and Fluit, A. C. (2004). Human-to-dog transmission of methicillin-resistant *Staphylococcus aureus*. *Emerg. Infect. Dis.*, *10*(**12**), 2235–7.

Van Heuverswyn, F., Li, Y., Bailes, E., Neel, C., Lafay, B., Keele, B. F., Shaw, K. S., Takehisa, J., Kraus, M. H., Loul, S., Butel, C., Liegeois, F., Yangda, B., Sharp, P. M., Mpoudi-Ngole, E., Delaporte, E., Hahn, B. H., and Peeters, M. (2007). Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. *Virology*, *368*(**1**), 155–71.

Van Loo, I., Huijsdens, X., Tiemersma, E., de Neeling, A., van de Sande-Bruinsma, N., Beaujean, D., Voss, A., and Kluytmans, J. (2007). Emergence of methicillin-resistant *Staphylococcus aureus* of animal origin in humans. *Emerg. Infect. Dis.*, *13*(**12**), 1834–1839.

Vancraeynest, D., Hermans, K., Martel, A., Vaneechoutte, M., Devriese, L. A., and Haesebrouck, F. (2004). Antimicrobial resistance and resistance genes in *Staphylococcus aureus* strains from rabbits. *Vet. Microbiol.*, *101*(**4**), 245–51.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.-E., and Etienne, J. (2003).

Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. *Emerg. Infect. Dis.*, *9*(**8**), 978–984.

Vasudevan, P., Nair, M. K. M., Annamalai, T., and Venkitanarayanan, K. S. (2003). Phenotypic and genotypic characterization of bovine mastitis isolates of *Staphylococcus aureus* for biofilm formation. *Vet. Microbiol.*, *92*(**1**), 179–185.

Verkaik, N. J., Benard, M., Boelens, H. A., de Vogel, C. P., Nouwen, J. L., Verbrugh, H. A., Melles, D. C., van Belkum, A., and van Wamel, W. J. B. (2010). Immune evasion cluster-positive bacteriophages are highly prevalent among human *Staphylococcus aureus* strains, but they are not essential in the first stages of nasal colonization. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.*, *17*(**3**), 343–348.

Viana, D., Blanco, J., Tormo-Más, M. Á., Selva, L., Guinane, C. M., Baselga, R., Corpa, J. M., Lasa, Í., Novick, R. P., Fitzgerald, J. R., and others. (2010). Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Mol. Microbiol.*, *77*(**6**), 1583–1594.

Vidal, P. M., Trindade, P. A., Garcia, T. O., Pacheco, R. L., Costa, S. F., Reinert, C., Hiramatsu, K., Mamizuka, E. M., Garcia, C. P., and Levin, A. S. (2009). Differences between "classical" risk factors for infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) and risk factors for nosocomial bloodstream infections caused by multiple clones of the staphylococcal cassette chromosome *mec* type IV. *Infect. Control Hosp. Epidemiol. Off. J. Soc. Hosp. Epidemiol. Am.*, *30*(**2**), 139–145.

VMD. (2012). UK Veterinary Antibiotic Resistance and Sales Surveillance., 1–104.

Von Eiff, C., Becker, K., Machka, K., Stammer, H., and Peters, G. (2001). Nasal carriage as a source of *Staphylococcus aureus* bacteraemia. *N. Engl. J. Med.*, *344*(**1**), 11.

Voss, A., Loeffen, F., Bakker, J., Klaassen, C., and Wulf, M. (2005). Methicillin-resistant *Staphylococcus aureu* in pig farming. *Emerg. Infect. Dis.*, *11*(**12**), 1965–1966.

Vybiral, D., Takãc, M., Loessner, M., Witte, A., Ahsen, U., and Bläsi, U. (2003). Complete nucleotide sequence and molecular characterization of two lytic *Staphylococcus aureus* phages: 44AHJD and P68. *FEMS Microbiol. Lett.*, *219*(**2**), 275–283.

Waage, S., Mørk, T., Røros, A., Aasland, D., Hunshamar, A., and Odegaard, S. A. (1999). Bacteria associated with clinical mastitis in dairy heifers. *J. Dairy Sci.*, *82*(**4**), 712–9.

Waldron, D. E., and Lindsay, J. A. (2006). Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *J. Bacteriol.*, *188*(**15**), 5578 –5585.

Walther, B., Monecke, S., Ruscher, C., Friedrich, A. W., Ehricht, R., Slickers, P., Soba, A., Wleklinski, C.-G., Wieler, L. H., and Lübke-Becker, A. (2009). Comparative molecular analysis substantiates zoonotic potential of equine methicillin-resistant *Staphylococcus aureus*. *J. Clin. Microbiol.*, *47*(**3**), 704–10.

Walther, B., Wieler, L. H., Friedrich, A. W., Hanssen, A.-M., Kohn, B., Brunnberg, L., and Lübke-Becker, A. (2008). Methicillin-resistant *Staphylococcus aureus* (MRSA) isolated from small and exotic animals at a university hospital during routine microbiological examinations. *Vet. Microbiol.*, *127*(**1-2**), 171–178.

Walther, B., Wieler, L. H., Vincze, S., Antão, E.-M., Brandenburg, A., Stamm, I., Kopp, P. A., Kohn, B., Semmler, T., and Lübke-Becker, A. (2012). MRSA variant in companion animals. *Emerg. Infect. Dis.*, *18*(**12**), 2017–20.

Wamel, W. J. B. van, Rooijakkers, S. H. M., Ruyken, M., Kessel, K. P. M. van, and Strijp, J. A. G. van. (2006). The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on β-Hemolysin-converting bacteriophages. *J. Bacteriol.*, *188*(**4**), 1310–1315.

Wang, S., Wu, C., Shen, J., Wu, Y., and Wang, Y. (2013). Hypermutable *Staphylococcus aureus* strains present at high frequency in subclinical bovine mastitis isolates are associated with the development of antibiotic resistance. *Vet. Microbiol.*, *165*(**3**), 410–415.

Wang, W.-Y., Chiueh, T.-S., Sun, J.-R., Tsao, S.-M., and Lu, J.-J. (2012). Molecular typing and phenotype characterization of methicillin-resistant *Staphylococcus aureus* isolates from blood in Taiwan. *PLoS One*, *7*(**1**), e30394.

Ward, P. N., Holden, M. T. G., Leigh, J. A., Lennard, N., Bignell, A., Barron, A., Clark, L., Quail, M. A., Woodward, J., Barrell, B. G., Egan, S. A., Field, T. R., Maskell, D., Kehoe, M., Dowson, C. G., Chanter, N., Whatmore, A. M., Bentley, S. D., and Parkhill, J. (2009). Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. *BMC Genomics*, *10*, 54.

Warr, G. W., Magor, K. E., and Higgins, D. A. (1995). IgY: clues to the origins of modern antibodies. *Immunol. Today*, *16*(**8**), 392–8.

Watanabe, S., Ito, T., Sasaki, T., Li, S., Uchiyama, I., Kishii, K., Kikuchi, K., Skov, R. L., and Hiramatsu, K. (2009). Genetic Diversity of Staphylocoagulase Genes (*coa*): Insight into the Evolution of Variable Chromosomal Virulence Factors in *Staphylococcus aureus*. *PLoS One*, *4*(**5**), e5714.

Wattinger, L., Stephan, R., Layer, F., and Johler, S. (2012). Comparison of *Staphylococcus aureus* isolates associated with food intoxication with isolates from human nasal carriers and human infections. *Eur. J. Clin. Microbiol. Infect. Dis.*, *31*(**4**), 455–64.

Weese, J. S., Archambault, M., Willey, B. M., Hearn, P., Kreiswirth, B. N., Said-Salim, B., McGeer, A., Likhoshvay, Y., Prescott, J. F., and Low, D. E. (2005). Methicillin-resistant *Staphylococcus aureus* in horses and horse personnel, 2000-2002. *Emerg. Infect. Dis.*, *11*(**3**), 430–435.

Weese, J. S., Rousseau, J., Traub-Dargatz, J. L., Willey, B. M., McGeer, A. J., and Low, D. E. (2005b). Community-associated methicillin-resistant *Staphylococcus aureus* in horses and humans who work with horses. *J. Am. Vet. Med. Assoc.*, *226*(**4**), 580–583.

Weese, J. S., Caldwell, F., Willey, B. M., Kreiswirth, B. N., McGeer, A., Rousseau, J., and Low, D. E. (2006). An outbreak of methicillin-resistant *Staphylococcus aureus* skin infections resulting from horse to human transmission in a veterinary hospital. *Vet. Microbiol.*, *114*(**1-2**), 160–164.

Weese, J. S., Dick, H., Willey, B. M., McGeer, A., Kreiswirth, B. N., Innis, B., and Low, D. E. (2006b). Suspected transmission of methicillin-resistant *Staphylococcus aureus* between domestic pets and humans in veterinary clinics and in the household. *Vet. Microbiol.*, *115*(**1-3**), 148–155.

Weese, J. S., Rousseau, J., Willey, B. M., Archambault, M., McGeer, A., and Low, D. E. (2006c). Methicillin-resistant *Staphylococcus aureus* in horses at a veterinary teaching hospital: frequency, characterization, and association with clinical disease. *J. Vet. Intern. Med.*, *20*(**1**), 182–186.

Wehrhahn, M. C., Robinson, J. O., Pearson, J. C., O'Brien, F. G., Tan, H. L., Coombs, G. W., Pascoe, E. M., Lee, R., Salvaris, P., Salvaris, R., New, D., and Murray, R. J. (2010). Clinical and laboratory features of invasive community-onset methicillin-resistant *Staphylococcus aureus* infection: a prospective case–control study. *Eur. J. Clin. Microbiol. Infect. Dis.*, *29*(**8**), 1025–1033.

Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G., Rose, D. J., Darling, A., Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C., and Blattner, F. R. (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* Serotype 2a strain 2457T. *Infect. Immun.*, *71*(**5**), 2775–2786.

Weigel, L. M. (2003). Genetic Analysis of a High-Level Vancomycin-Resistant isolate of *Staphylococcus aureus*. *Science (80-.).*, *302*(**5650**), 1569–1571.

Weinert, L. A., Welch, J. J., Suchard, M. A., Lemey, P., Rambaut, A., and Fitzgerald, J. R. (2012). Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. *Biol. Lett.*, *8*(**5**), 829–832.

Wentworth, B. B. (1963). Bacteriophage typing of the Staphylococci. *Bacteriol. Rev.*, *27*, 253–72.

Westh, H., Zinn, C. S., and Rosdahl, V. T. (2004). An international multicenter study of antimicrobial consumption and resistance in *Staphylococcus aureus* isolates from 15 hospitals in 14 countries. *Microb. Drug Resist.*, *10*(**2**), 169–76.

Wieler, L. H., Ewers, C., Guenther, S., Walther, B., and Lübke-Becker, A. (2011). Methicillin-resistant staphylococci (MRS) and extended-spectrum beta-lactamases (ESBL)-producing *Enterobacteriaceae* in companion animals: Nosocomial infections as one reason for the rising prevalence of these potential zoonotic pathogens in clinical samples. *Int. J. Med. Microbiol.*, *301*(**8**), 635–641.

Wilder-Smith, A., Leong, H. N., and Villacian, J. S. (2006). In-flight transmission of Severe Acute Respiratory Syndrome (SARS): A Case Report. *J. Travel Med.*, *10*(**5**), 299–300.

Williams, R. E. O. (1963). Healthy carriage of *Staphylococcus aureus*: its prevalence and importance. *Microbiol. Mol. Biol. Rev.*, *27*(**1**), 56–71.

Wilson, G. J., Seo, K. S., Cartwright, R. A., Connelley, T., Chuang-Smith, O. N., Merriman, J. A., Guinane, C. M., Park, J. Y., Bohach, G. A., Schlievert, P. M., Morrison, W. I., and Fitzgerald, J. R. (2011). A novel core genome-encoded superantigen contributes to lethality of community-associated MRSA necrotizing pneumonia. *PLoS Pathog*, *7*(**10**), e1002271.

Wolf, C., Kusch, H., Monecke, S., Albrecht, D., Holtfreter, S., von Eiff, C., Petzl, W., Rainard, P., Bröker, B. M., and Engelmann, S. (2011). Genomic and proteomic characterization of *Staphylococcus aureus* mastitis isolates of bovine origin. *Proteomics*, *11*(**12**), 2491–2502.

Wolfe, N. D., Dunavan, C. P., and Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, *447*(**7142**), 279–83.

Woolhouse, M., and Gaunt, E. (2007). Ecological origins of novel human pathogens. *Crit. Rev. Microbiol.*, *33*(**4**), 231–242.

Wulf, M. W. H., Markestein, A., van der Linden, F. T., Voss, A., Klaassen, C., and Verduin, C. M. (2008). First outbreak of methicillin-resistant *Staphylococcus aureus* ST398 in a Dutch hospital, June 2007. *Eur. Commun. Dis. Bull.*, *13*(**9**), 1854–1861.

Wyllie, D. H., Walker, A. S., Miller, R., Moore, C., Williamson, S. R., Schlackow, I., Finney, J. M., O'Connor, L., Peto, T. E. A., and Crook, D. W. (2011). Decline of meticillin-resistant *Staphylococcus aureus* in Oxfordshire hospitals is strain-specific and preceded infection-control intensification. *BMJ Open*, *1*(**1**), e000160.

Xu, B. L., Zhang, G., Ye, H. F., Feil, E. J., Chen, G. R., Zhou, X. M., Zhan, X. M., Chen, S. M., and Pan, W. B. (2009). Predominance of the Hungarian clone (ST 239-III) among hospital-acquired meticillin-resistant *Staphylococcus aureus* isolates recovered throughout mainland China. *J. Hosp. Infect.*, *71*(**3**), 245–55.

Xu, S.-Y., Corvaglia, A. R., Chan, S.-H., Zheng, Y., and Linder, P. (2011). A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. *aureus* USA300. *Nucleic Acids Res.*, *39*(**13**), 5597–610.

Yamada, T., Tochimaru, N., Nakasuji, S., Hata, E., Kobayashi, H., Eguchi, M., Kaneko, J., Kamio, Y., Kaidoh, T., and Takeuchi, S. (2005). Leukotoxin family genes in *Staphylococcus aureus* isolated from domestic animals and prevalence of *lukM-lukF-PV* genes by bacteriophages in bovine isolates. *Vet. Microbiol.*, *110*(**1-2**), 97–103.

Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., Yamada, S., Komatsuzawa, H., and Sugai, M. (2000). Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. *Mol. Microbiol.*, *38*(**4**), 694–705.

Yoshida, H., Bogaki, M., Nakamura, S., Ubukata, K., and Konno, M. (1990). Nucleotide sequence and characterization of the *Staphylococcus aureus norA* gene, which confers resistance to quinolones. *J. Bacteriol.*, *172*(**12**), 6942–6949.

Young, B., and Golubchik, T. (2012). Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc. Natl. Acad. Sci.*, *109*(**12**), 4550–4555.

Zadoks, R., Van Leeuwen, W., Barkema, H., Sampimon, O., Verbrugh, H., Schukken, Y. H., and Van Belkum, A. (2000). Application of pulsed-field gel electrophoresis and binary typing as tools in veterinary clinical microbiology and molecular epidemiologic analysis of bovine and human *Staphylococcus aureus* isolates. *J. Clin. Microbiol.*, *38*(**5**), 1931–1939.

Zecconi, A., Cesaris, L., Liandris, E., Daprà, V., and Piccinini, R. (2006). Role of several *Staphylococcus aureus* virulence factors on the inflammatory response in bovine mammary gland. *Microb. Pathog.*, *40*(**4**), 177–183.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U. S. A.*, *105*(**33**), 11597–604.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.*, *18*(**5**), 821–829.

Zhang, K., McClure, J.-A., Elsayed, S., and Conly, J. M. (2009). Novel staphylococcal cassette chromosome *mec* type, tentatively designated type VIII, harboring class A *mec* and type 4 *ccr* gene complexes in a Canadian epidemic strain of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, *53*(**2**), 531–540.

Zinn, C. S., Westh, H., and Rosdahl, V. T. (2004). An international multicenter study of antimicrobial resistance and typing of hospital *Staphylococcus aureus* isolates from 21 laboratories in 19 countries or states. *Microb. Drug Resist.*, *10*(**2**), 160–168.

Zong, Y., Xu, Y., Liang, X., Keene, D. R., Höök, A., Gurusiddappa, S., Höök, M., and Narayana, S. V. L. (2005). A "Collagen Hug" model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J.*, *24*(**24**), 4224–36.

Appendices

Appendix 1: Additional phylogenetic trees

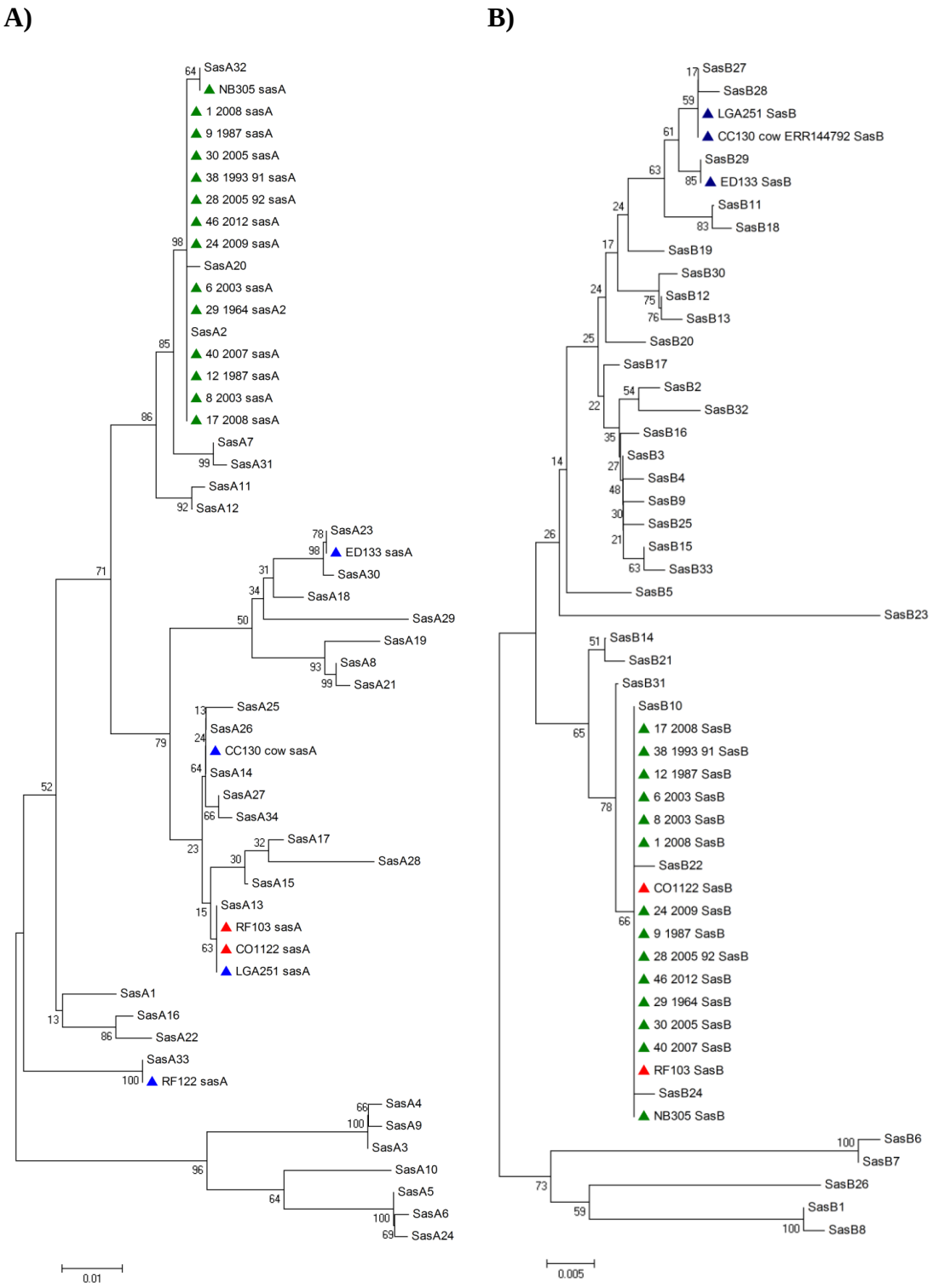


Figure S.1.A) Neighbour joining tree for *sasA* alleles. B) Neighbour joining tree for *sasB* alleles. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, as indicated at each node.

C) **D)**

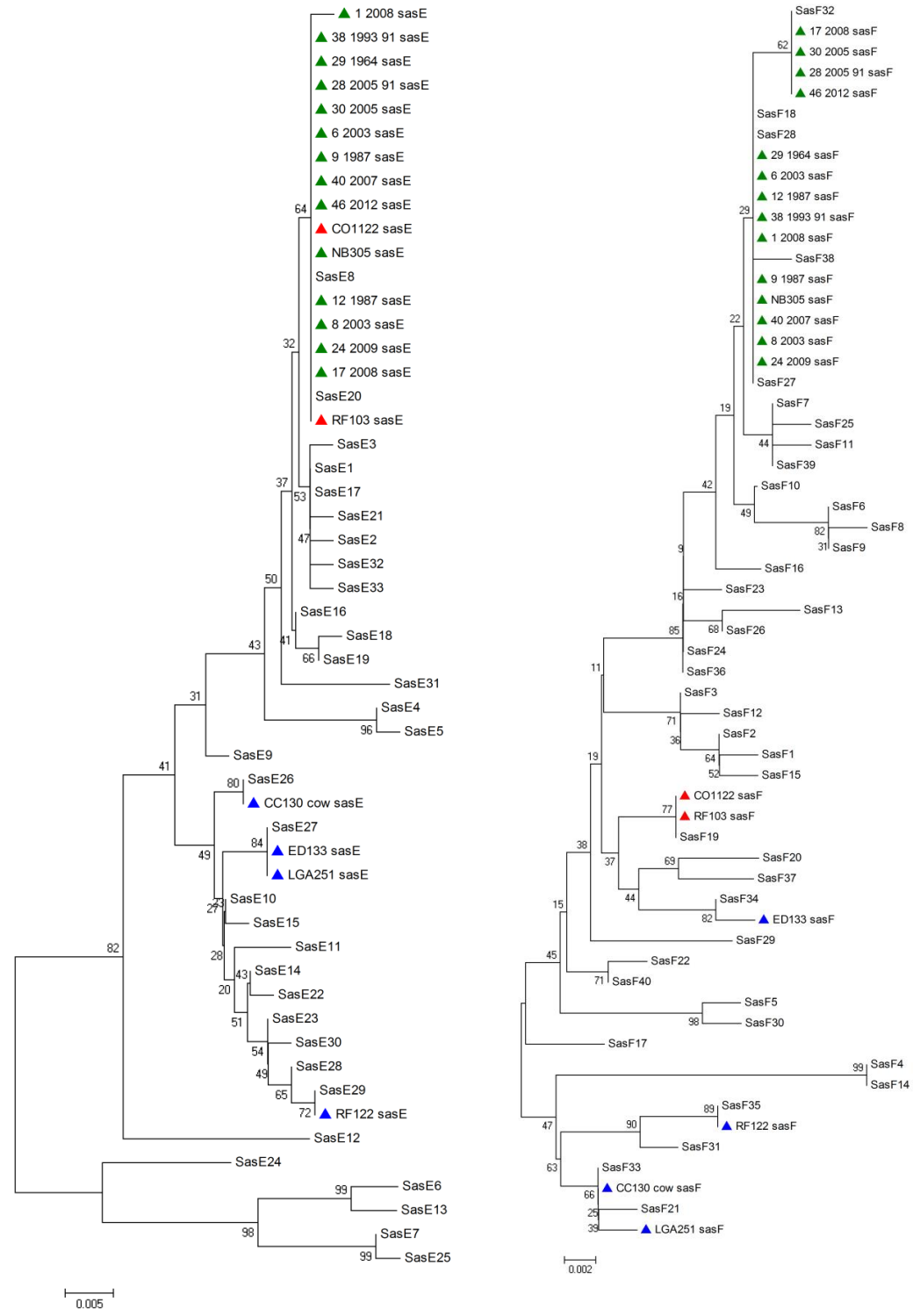


Figure S.1.C) Neighbour joining tree for *sasE* alleles. D) Neighbour joining tree for *sasF* alleles. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, as indicated at each node.

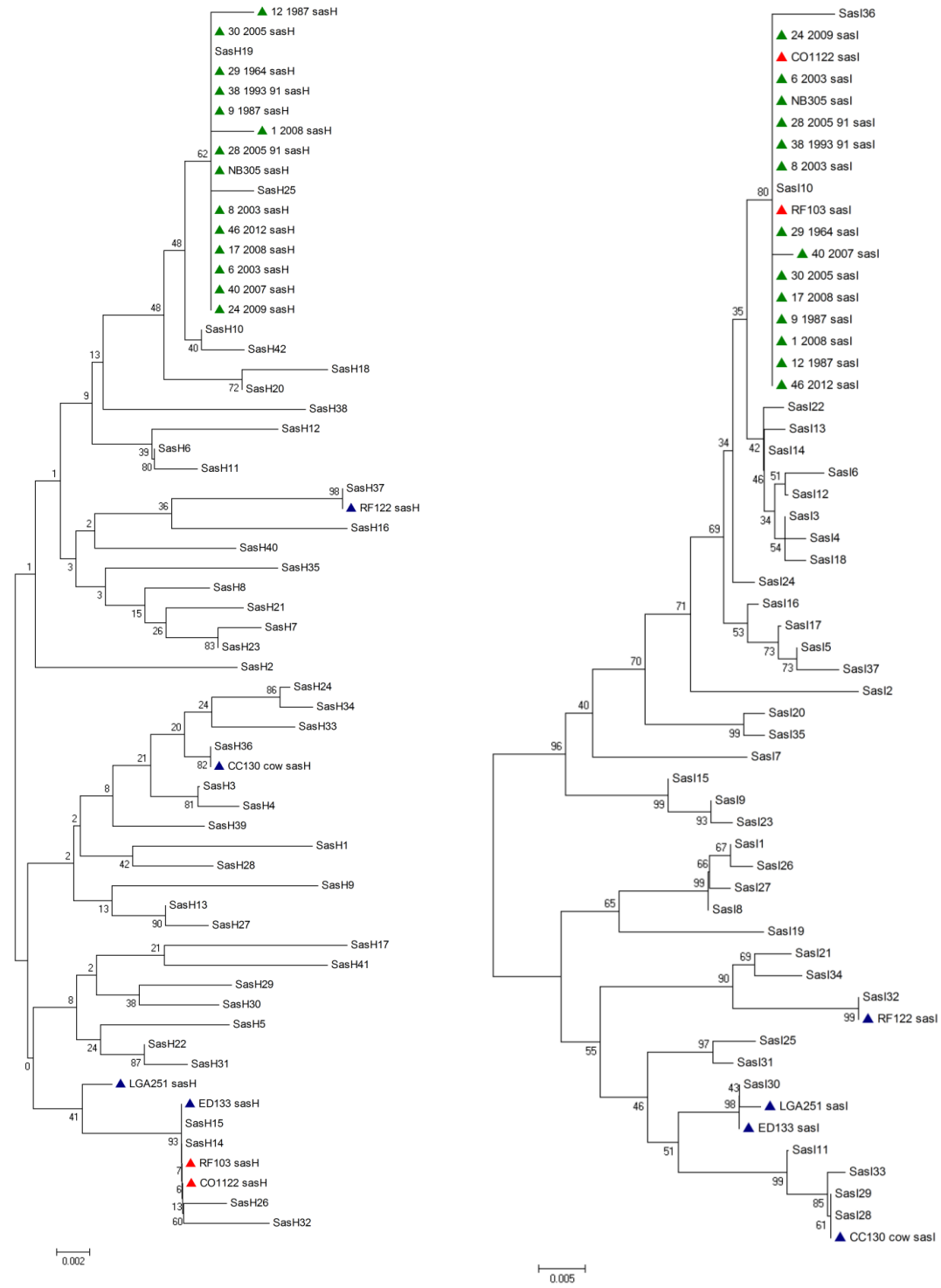
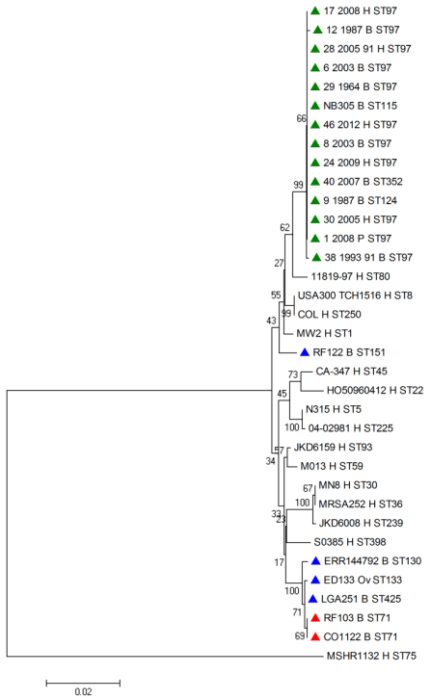
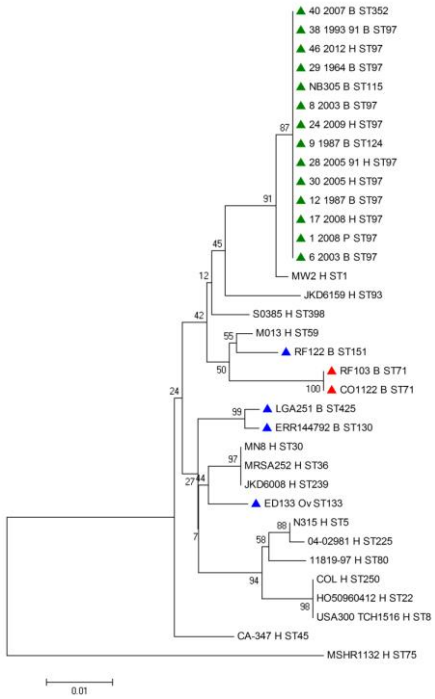


Figure S.1.E) Neighbour joining tree for *sasH* alleles. F) Neighbour joining tree for *sasI* alleles. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node.

A) MW2547 (*phoB*-alkaline phosphatise III precursor)



B) MW2611 (Hypothetical protein)



C) MW2630 (*trmE*-tRNA modification GTPase)

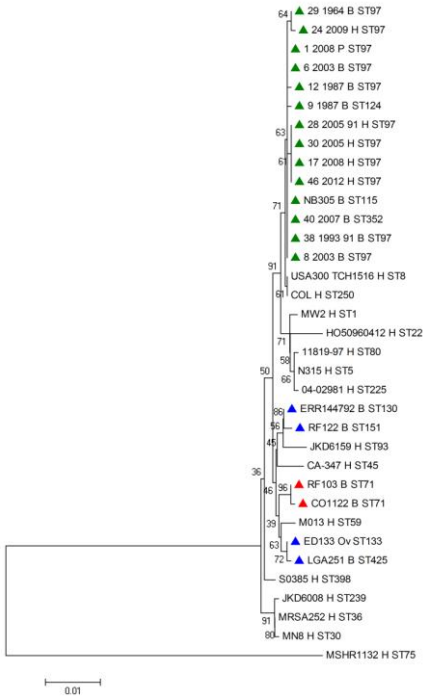
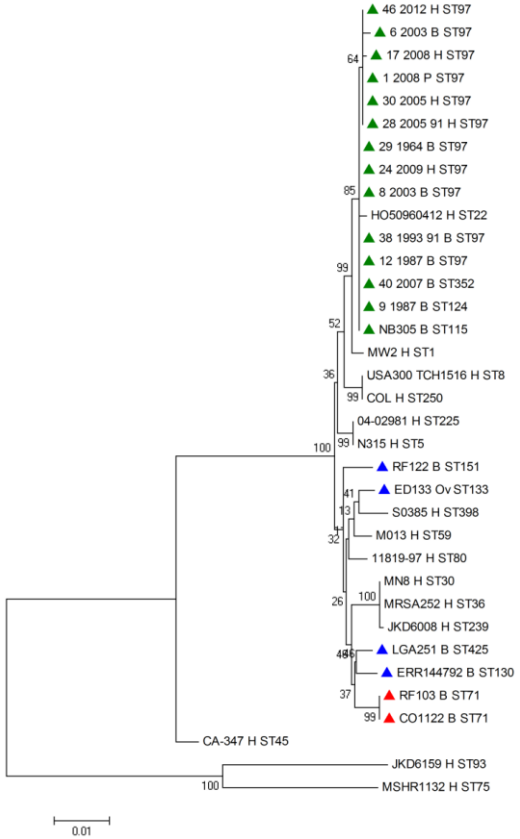


Figure S.2 (A-C). Neighbour joining trees for CDS within "SNP dense" region. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

D) MW0008 (*hutU*-histidine ammonia-lyase)



E) MW0016 (*dnaC*-replicative DNA helicase)

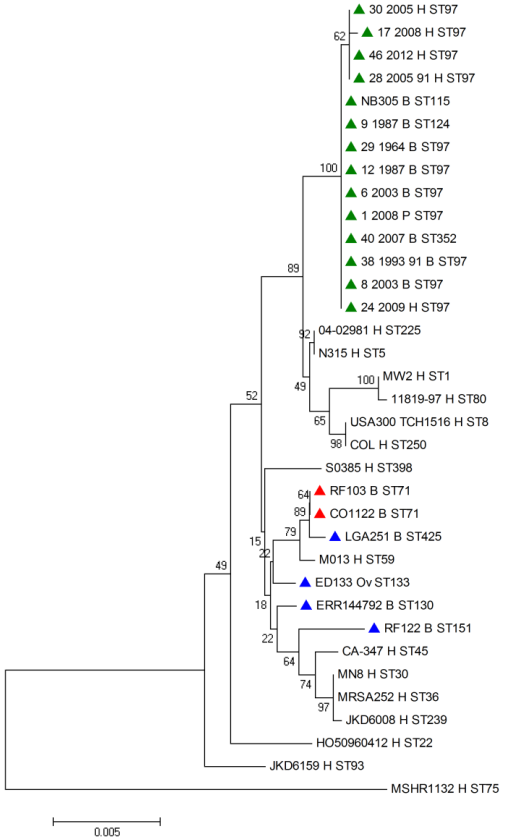
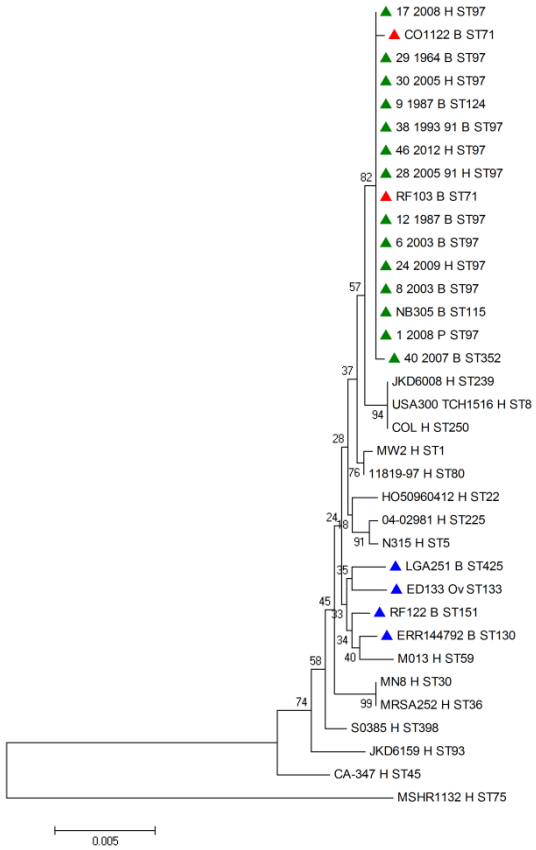
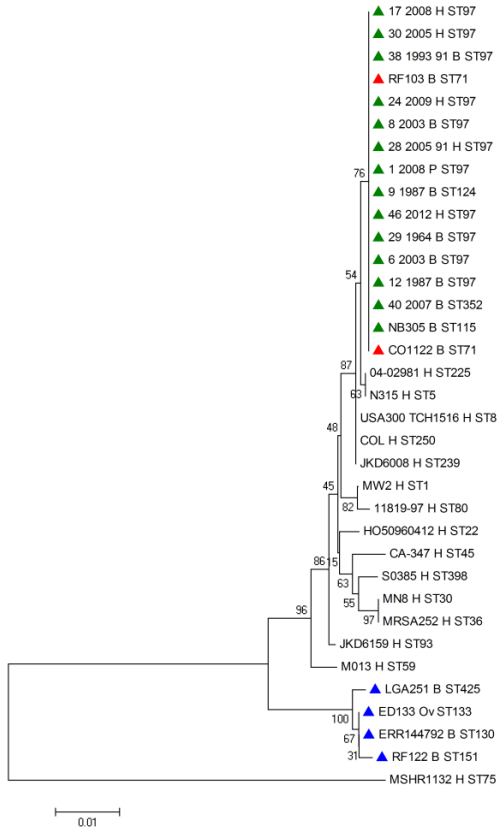


Figure S.2 (A,B). Neighbour joining trees for CDS within "SNP dense" region. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

A) MW0365 (*pbuX*-xanthin permease)



B) MW0626 (hypothetical protein)



C) MW1183(*glpK*-glycerol kinase)

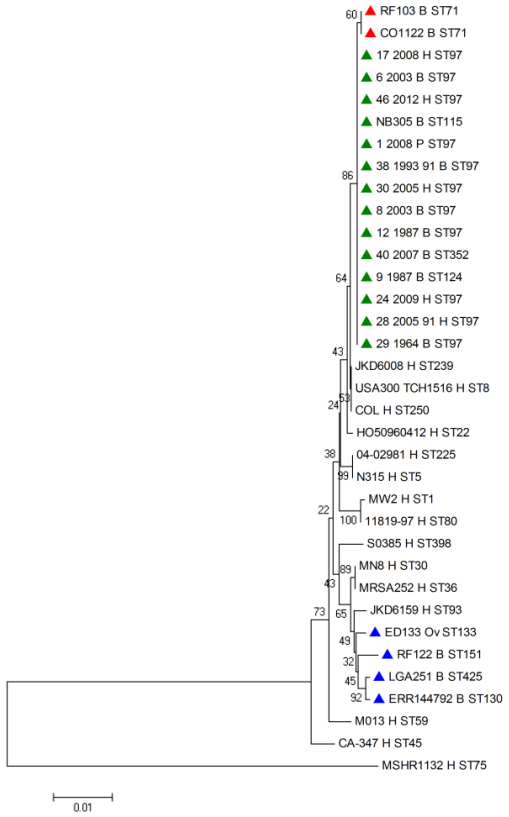
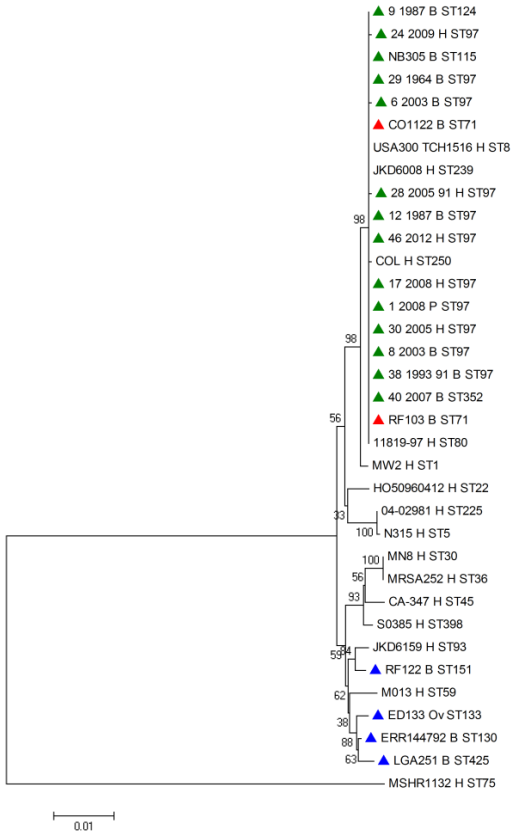


Figure S.3 (A-C). Neighbour joining trees for CDS outwith "SNP dense" region. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

D) MW1310 (*ctpA*-carboxy-terminal processing proteinase)



E) MW2252 (*hutU*-urocanate hydratase)

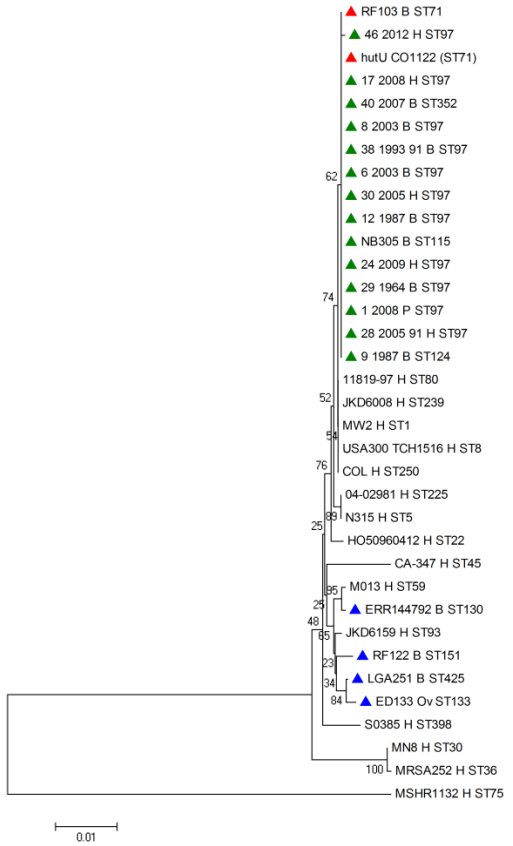


Figure S.3 (D,E). Neighbour joining trees for CDS outwith "SNP dense" region. Strains are colour-coded according to genotype (red=ST71, green=CC97, blue=non-CC97). Both phylogenetic trees produced using MEGA v4.0 with 1000 bootstrap replicates, indicated at each node. Taxa are individually labelled according to strain identity, host (B=bovine, H=human, P=pig) and MLST sequence type.

Appendix 2: Published Paper